Maladies chroniques et traumatismes

Système multi sources de surveillance des cancers (SMSC)

Etude pilote sur le cancer de la thyroïde en Ile-de-France et dans le Nord-Pas-de-Calais, 2010

Rapport final d'étude
Octobre 2010

INSTITUT DE VEILLE SANITAIRE

Sommaire

Abréviations 2

Liste des tableaux et figures 5

Résumé 6

Préambule 8

1. Historique, bases et principes du SMSC 10

1.1 Historique du système national de surveillance multi sources des cancers (SMSC) 10

1.2 Les bases du SMSC (justification, principe, objectif) 12

2. Préparation de l'étude pilote du SMSC 16

2.1 Les objectifs de l'étude pilote 16

2.2 Périmètre de l'étude pilote 16

2.3 Définition du schéma d'architecture d'anonymisation 23

2.4 Méthodologie 26

3. Réalisation de l'étude pilote 30

3.1 Les travaux administratifs 30

3.2 Les travaux techniques : le développement de la solution informatique SMSC 33

3.3 Les travaux institutionnels connexes 39

4. Résultats de l'analyse des données croisées 40

4.1 Analyse des cas rétrospectifs de 2004 40

4.2 Transmission des fiches ACP de l'année 2009 pour le volet prospectif du pilote 51

4.3 Discussion 51

5. Conclusion et perspectives 53

5.1 Malgré les difficultés, le pilote a atteint ses objectifs d'étude de faisabilité 53

5.2 Une évolution du contexte autour des systèmes d'information partagée 54

5.3 Les perspectives 55

Références bibliographiques 57

Annexes 59

Système multi sources de surveillance des cancers (SMSC)

Etude pilote sur le cancer de la thyroïde en Ile-de-France et dans le Nord-Pas-de-Calais, 2010

Rapport final d'étude

Octobre 2010

Auteurs :

Institut de veille sanitaire (InVS), Département des maladies chroniques et traumatismes (DMCT)

Céline CASERIO-SCHÖNEMANN
Yao KUDJAWU
Laurence CHÉRIÉ-CHALLINE
Juliette BLOCH

InVS, Service des systèmes d'information (SSI)

Agnès GUILLET
Arnaud MUSSET

InVS, Direction scientifique (DS)

Javier NICOLAU

Remerciements à l'ensemble des partenaires du projet, en particulier :

- la Caisse nationale de l'assurance maladie des travailleurs salariés (CnamTS) ;
- le syndicat des anatomo-pathologistes ;
- les médecins pathologistes des structures d'anatomo cyto pathologie (ACP) impliquées ;
- les Directions régionales des affaires sociales et sanitaires de l'Ile-de-France et du Nord-Pas-de-Calais ;
- le Centre de regroupement informatique et statistique en anatomie et pathologie d'Ile-de-France (Crisapif) ;
- l'Institut national pour la santé et la recherche médicale (Inserm).

Abréviations

ACP	Anatomie et cytologie pathologiques
Adeli	Automatisation des listes
Afaqap	Association française d'assurance qualité en anatomie pathologie
ALD(30)	Affection de longue durée (30 pathologies)
Amoa	Maîtrise d'ouvrage
ARS	Agence régionale de santé
ASIP-Santé	Agence des systèmes d'information partagés en santé
ATIH	Agence technique de l'information hospitalière
BA	Bulletin d'adhésion
Camel	Chiffrement assurance maladie élargie
CCAM	Classification commune des actes médicaux
CdAM	Catalogue des actes médicaux
Centi	Centre national de traitement de l'information
CépiDc	Centre d'épidémiologie sur les causes médicales de décès
Cessi	Centre d'étude sur la sécurité des systèmes d'information
CHR	Centre hospitalier régional
CHRU	Centre hospitalier régional et universitaire
CIM10	Classification internationale des maladies, 10e version
Cire	Cellule inter régionale d'épidémiologie
CLCC	Centre de lutte contre le cancer
CnamTS	Caisse nationale d'assurance maladie des travailleurs salariés
Cnil	Commission nationale de l'informatique et des libertés
CNR	Comité national des registres
CNSS	Commission nationale de sécurité des systèmes d'information
CoGes	Comité de gestion
CoStrat	Comité stratégique
CRFS	Compte rendu fiche standardisé
Crisap	Centre de regroupement informatique et statistique en anatomie et pathologie
Crisapif	Centre de regroupement informatique et statistique en anatomie et pathologie d'Ile-de-France
CSN	Centre de support national

DCC	Dossier communiquant cancer
DIM	Département de l'information médicale
DMCT	Département des maladies chroniques et traumatismes
DMP	Dossier médical personnel
DO	Déclaration obligatoire
Drass	Direction régionale des affaires sanitaires et sociales
ETP	Equivalent temps plein
FCRISAP	Fédération des centres de regroupement informatique et statistique en anatomie et pathologie
Foin	Fonction d'occultation d'identifiants nominatifs
Francim	France incidence mortalité (réseau des registres français des cancers)
GHM	Groupe homogène de malades
GMSIH	Groupement pour la modernisation du système d'information hospitalier
GT	Groupe de travail
HL7 CDA-R2	Health Level Seven Clinical Document Architecture Release 2
IdF	Ile-de-France
Inserm	Institut national pour la santé et la recherche médicale
IFR	Institut fédératif de recherche
InVS	Institut de veille sanitaire
Inca	Institut national du cancer
MOA	Maîtrise d'ouvrage
MOE	Maître d'œuvre
MSA	Mutualité sociale agricole
NPdC	Nord-Pas-de-Calais
NIR	Numéro d'inscription au répertoire
PMSI	Programme de médicalisation des systèmes d'information
PSPH	Etablissement de santé privé participant au service public hospitalier
RRC	Réseaux régionaux de cancérologie
Retex	Retour d'expérience
Rida	Relevé d'information décision action
RSA	Résumé de sortie anonymisé
RSI	Régime social des indépendants

SDSI	Schéma directeur des systèmes d'information
SGL	Système de gestion de laboratoire
SIH	Système d'information hospitalier
SMPF	Syndicat des médecins pathologistes français
SMSC	Système multi sources cancer
Sniiram	Système national d'information interrégimes de l'Assurance maladie
SNC	Système nerveux central
SSI	Service des systèmes d'information
VSR	Vérification de service régulier

LISTE DES TABLEAUX

Tableau 1 : Editeurs de logiciels ACP – Représentation au niveau national en 2005, étude pilote SMSC, 2004-2010… 17
Tableau 2 : Structures ACP pilotes et secteur d'activité, étude pilote SMSC, 2004-2010…….. 18
Tableau 3 : Les informations de données ACP recueillies……... 21
Tableau 4 : Les informations ALD recueillies…….. 22
Tableau 5 : Synthèse des principaux jalons……. 30
Tableau 6 : Récapitulatif du cadre contractuel, étude pilote SMSC, 2004-2010……... 32
Tableau 7 : Première consultation marché SMSC – propositions financières des éditeurs de logiciels ACP, étude pilote SMSC, 2004-2010……... 33
Tableau 8 : Bilan par site de la transmission agrégée des données ACP 2004……... 40
Tableau 9 : Pourcentage d'informations recueillies et exploitables dans la base ACP, 2004 et description des cas…... 41
Tableau 10 : Origine des cas de cancer de la thyroïde mis en ALD appariés avec les données du PMSI ou des ACP… 43
Tableau 11 : Résultat du croisement des données de la base ACP avec les données PMSI et ALD30……. 43
Tableau 12 : Composition de la base d'analyse…….. 43
Tableau 13 : Niveau de cohérence des données entre le PMSI et les ALD…….. 45
Tableau 14 : Niveau de cohérence entre les données croisées (PMSI/ALD/ACP)+……... 45
Tableau 15 : Niveau de cohérence entre les données ACP, PMSI et ALD croisées deux à deux……. 46
Tableau 16 : Effectif et taux d'appariement selon le type de croisement…… 46
Tableau 17 : Caractéristiques des cas absents du PMSI…… 48
Tableau 18 : Caractéristiques des cas absents des ALD…… 49
Tableau 19 : Caractéristiques des cas ACP présents dans la base ALD30 ou PMSI et absents de la base ALD30 et de la base PMSI……... 50
Tableau 20 : Répartition des cas de cancer de la thyroïde tiré au sort par structure ACP et résultats de l'exercice de retour au cas……... 51

LISTE DES FIGURES

Figure 1 : Carte présentant la situation des registres du cancer en France métropolitaine qualifiés par le CNR au 1er janvier 2010……..... 10
Figure 2 : Partenaires et flux de données, étude pilote SMSC, 2004-2010…….. 20
Figure 3 : Circuits simplifiés d'anonymisation, étude pilote SMSC, 2004-2010……... 24
Figure 4 : Problématique d'interconnexion réseau, étude pilote SMSC, 2004-2010……. 36
Figure 5 : Frise chronologique de la prestation SMSC…… 38
Figure 6 : Répartition des effectifs appariés selon les sources…… 44
Figure 7 : Comparaison des taux d'appariement de la région du Nord-Pas-de-Calais avec ceux du registre de l'agglomération de Lille de 2005…… 47
Figure 8 : Architecture des flux DCC, étude pilote SMSC, 2004-2010 …….. 55

RÉSUMÉ

1 – Contexte

Le système de surveillance des cancers en France repose principalement sur les registres des cancers qui ne couvraient que 13 % de la population en 2000 (20 % aujourd'hui). La réflexion sur le renforcement du dispositif de surveillance des cancers en France est issue des travaux réalisés par la commission multidisciplinaire mise en place et animée par l'InVS en 2000 dans le cadre de la mission d'évaluation des conséquences sanitaires en France de l'accident de Tchernobyl. Le développement d'un système de surveillance nationale, le système multi sources cancer (SMSC), en complément du système reposant sur les registres a été recommandé par la commission et repris dans le premier plan cancer gouvernemental 2003-2007.

Le principe du SMSC est d'appliquer sur tout le territoire, dans le cadre d'un système automatisé, passif et anonymisé, la démarche utilisée par les registres de cancer pour l'enregistrement des cas incidents de cancer. Cet enregistrement repose sur le croisement de plusieurs sources d'information. Toutefois, à la différence du schéma envisagé pour le SMSC, les registres travaillent au niveau local, avec recherche active des informations (retour au dossier) et sur une base nominative.

Le SMSC repose sur le croisement de trois sources à savoir :
- les données issues du Programme de médicalisation des systèmes d'information (PMSI) ;
- les données de l'Assurance maladie relatives aux exonérations du ticket modérateur pour cancer, correspondant à l'affection longue durée n°30 ou ALD cancer ;
- les données issues des comptes-rendus d'ACP permettant d'attester du diagnostic de cancer (hors hémopathies malignes).

Les données des deux premières bases sont centralisées au niveau national tandis que la base ACP est à construire.
L'objectif de l'étude pilote était d'étudier la pertinence et la faisabilité de la mise en place du SMSC.

2- Matériels et méthode

L'étude pilote a porté sur le cancer de la thyroïde et s'est déroulée dans deux régions, l'Ile-de-France (IdF) et le Nord-Pas-de-Calais (NPdC). Elle a impliqué un réseau de partenaires qui avaient pour tâches :

- de rassembler, d'anonymiser, et de transmettre les comptes-rendus ACP rétrospectifs de cancer de la thyroïde diagnostiqué en 2004 dans les huit structures ACP volontaires chez des patients résidant dans les deux régions pilotes ;
- ou de jouer le rôle de tiers de confiance (alerte pour le retour au cas).

Le recueil des données a été réalisé grâce à un formulaire informatisé installé sur des postes dédiés dans chacune des huit structures ACP.
Les partenaires de l'étude pilote étaient situés à trois niveaux géographiques : local (8 structures ACP volontaires), régional (le Centre de regroupement en informatique et statistique en ACP d'Ile-de-France et les Directions régionales des affaires sanitaires et sociales d'Ile-de-France et du Nord-Pas-de-Calais devenues Agences régionales de santé) puis national (la Caisse nationale d'assurance maladie des travailleurs salariés et l'Inserm IFR69).
Les données individuelles d'hospitalisation et de mise en ALD 2004 pour cancer thyroïdien des patients domiciliés en IDF ou dans le NPdC ont été respectivement extraites du PMSI et du Système national d'information inter régime de l'Assurance maladie (Sniiram).
La mise en place du dispositif, les difficultés rencontrées et les données recueillies sont décrites dans le rapport.
Un test de retour au cas a été effectué.

3- Résultats

Au total, 423 fiches de données ACP de cancer de la thyroïde transmises par les ACP des deux régions pilotes ont été validées par l'InVS. Ces données ont été croisées avec celles des ALD cancer (1 159 bénéficiaires) et du PMSI (2 947 patients) de l'année 2004.
Le taux d'appariement (nombre de cas ACP retrouvés dans les autres sources) observé était de 84 % entre les trois sources, de 81 % entre les sources PMSI et ACP et de 48 % entre les sources ACP et ALD cancer.
On notait une concordance totale sur le sexe, l'âge et le département de domiciliation pour les données appariées.
A l'issue du test de retour au cas, la totalité des dossiers transmis a été retrouvée.
La mise en place du système a mis en évidence que la solution informatique retenue atteignait un niveau de complexité tel qu'il n'était pas envisageable de la déployer en l'état au niveau national. Cette complexité est liée aux difficultés techniques rencontrées, principalement :

- les contraintes liées aux outils d'anonymisation et de sécurisation des données ;
- l'adaptation en fonction des sites de la solution informatique de remontée des données ;
- l'hétérogénéité des systèmes d'information hospitaliers ;
- l'instabilité des flux dans la transmission des données responsable de ruptures récurrentes sur toute la chaîne de remontée des données.

4- Conclusion et perspectives

Au regard de ces difficultés liées essentiellement à la mise en place de la base nationale des données ACP, et compte-tenu de l'évolution du contexte de développement des systèmes d'information en santé (création de l'Agence des systèmes d'information partagés en santé), deux axes de développement du SMSC au niveau national ont été retenus :

1- poursuivre dans l'immédiat au niveau national, les travaux autour du croisement des deux sources PMSI et ALD cancer déjà disponibles, à des fins descriptives et de production d'indicateurs de surveillance des cancers ;

2- à moyen terme et en parallèle avec le premier axe, étudier les conditions de mise en place de la base nationale de données ACP. La constitution de cette base reposera sur le dossier communiquant de cancérologie (DCC) dont un des enjeux principaux sera d'utiliser un identifiant national qui soit commun aux sources PMSI et ALD cancer afin de permettre leur croisement.

PRÉAMBULE

Les missions de l'Institut de veille sanitaire (InVS) ont été définies par la loi n° 98-535 du 1er juillet 1998 relative au renforcement de la veille sanitaire et du contrôle de la sécurité sanitaire des produits destinés à l'homme.
Elles recouvrent :

- **la surveillance et l'observation permanentes de l'état de santé de la population** ; l'InVS participe au recueil et au traitement des données sur l'état de santé de la population à des fins épidémiologiques, en s'appuyant notamment sur des correspondants publics et privés constituant le Réseau national de santé publique ;
- **la veille et la vigilance sanitaires** ; l'InVS est chargé de rassembler, analyser et actualiser les connaissances sur les risques sanitaires, leurs causes et leur évolution ; de détecter de manière prospective les facteurs de risque susceptibles de modifier ou d'altérer la santé de la population ou de certaines de ses composantes, de manière soudaine ou diffuse ; d'étudier et de répertorier, pour chaque type de risque, les populations les plus fragiles ou menacées ;
- **l'alerte sanitaire** ; l'InVS doit informer sans délai le ministre chargé de la Santé en cas de menace pour la santé de la population ou de certaines de ses composantes, quelle qu'en soit l'origine, et lui recommander toute mesure ou action appropriée pour prévenir la réalisation ou atténuer l'impact de cette menace ;
- **une contribution à la gestion des situations de crise sanitaire** ; l'InVS propose aux pouvoirs publics toute mesure ou action nécessaire.

L'InVS participe à l'action européenne et internationale de la France, notamment à des réseaux internationaux de santé publique dédiés à la surveillance et à l'alerte sanitaires.

Les missions de l'InVS ont été réaffirmées et renforcées, notamment dans la dimension de l'alerte, par la loi du 9 août 2004 relative à la politique de santé publique, ainsi que par la loi du 24 juin 2009 - Hôpital patients santé territoires.

Dans le domaine du cancer, ses missions, développées depuis 1999 par le Département des maladies chroniques et traumatismes (DMCT) en lien avec les autres départements scientifiques, sont les suivantes :

- la surveillance épidémiologique des cancers, avec la production des indicateurs de surveillance définis notamment pour le suivi du plan cancer. Elle constitue une aide pour les décideurs dans le pilotage et l'évaluation des mesures de prévention et de prise en charge des cancers. Elle est assurée par un réseau de partenaires de l'InVS qui comprend notamment les registres du cancer ;
- l'alerte des pouvoirs publics en cas de menace pour la santé. Cette mission d'alerte s'exerce en cas de signalement, par la population ou les autorités sanitaires, d'agrégats spatio-temporels ou clusters de cancers (regroupement de cas dans le temps et à un niveau géographique limité). Dans ce cas, les cellules interrégionales d'épidémiologie (Cire) ont pour mission de mener les investigations nécessaires afin d'objectiver un éventuel excès de cas, en relation ou non avec une exposition commune environnementale ou professionnelle ;
- l'évaluation des programmes de dépistage organisé des cancers du sein, du col utérin et colorectaux.

La lutte contre le cancer est une priorité nationale. Malgré les progrès réels réalisés dans les domaines de la prévention, du dépistage et du traitement, les cancers restent un problème majeur de santé publique en France, comme dans l'ensemble des pays développés. Ils restent la première cause de mortalité devant les pathologies cardio-vasculaires.
Les projections pour 2010 de l'incidence et de la mortalité par cancer en France métropolitaine, produits conjointement par le réseau français des registres des cancers (Francim), le service de biostatistique des Hospices civils de Lyon (HCL), l'InVS et l'Institut national du cancer (INCa), dans le cadre de leur partenariat, font état d'une estimation de 357 500 nouveaux cas de cancer (203 000 chez l'homme et 154 500 chez la femme) et de 146 500 nouveaux décès (84 500 chez l'homme et 62 000 chez la femme) [1].
L'évolution des cancers ces 25 dernières années en France a été marquée par des divergences entre l'incidence et de la mortalité. Alors que l'incidence a considérablement augmenté, le risque de mortalité par cancer a diminué. Cette divergence entre mortalité et incidence s'explique par l'évolution croisée des cancers : les tumeurs les plus agressives (œsophage, estomac, voies aérodigestives supérieures) ont chuté ces dernières années chez l'homme en lien avec la diminution de la consommation alcoolo-tabagique tandis que les cancers de pronostic plus favorables, pouvant être diagnostiqués très précocement, ont augmenté (sein, prostate) [2].
Face à ces données, le chef de l'État a érigé la lutte contre le cancer comme priorité nationale et a lancé en 2004 le premier plan quinquennal gouvernemental, auquel a fait suite un second plan couvrant la période 2009-2013 [3]. Le premier plan cancer confie à l'InVS la mission de renforcer la surveillance des cancers notamment au travers de deux premières mesures :

- mesure 1 – Soutien des registres des cancers et développement du système national de l'InVS ;
- mesure 2 – Développement par l'InVS des analyses régionales et assistance aux politiques régionales de santé.

La demande des pouvoirs publics de développer un système de surveillance nationale fait suite aux travaux de l'InVS sur les conséquences sanitaires de l'accident de Tchernobyl et sur la pertinence du dispositif de surveillance des cancers existant. Cette réflexion dépassait le cadre *stricto sensu* de l'accident de Tchernobyl pour aborder la question plus générale des risques environnementaux pouvant survenir en tout point du territoire (source potentiellement polluante – usines d'incinération d'ordures ménagères, centrales nucléaires – ou accident généralisé) [3;4]. Le développement d'un Système multisources national de surveillance des cancers (SMSC) qui viendrait compléter le dispositif de couverture partielle du territoire reposant sur les registres des cancers a été recommandé dans ce cadre.
Afin d'évaluer la faisabilité d'un tel système, une étude pilote a été conduite par l'InVS entre 2004 et 2010 sur le cancer de la thyroïde en Ile-de-France et dans le Nord-Pas-de-Calais.

Ce rapport a pour objectifs de :

- rappeler les fondements historiques et épidémiologiques ayant conduit à l'élaboration du projet de Système multi - sources de surveillance des cancers (SMSC) ;
- décrire les grands principes du SMSC ;
- détailler la réalisation, dans toutes ses dimensions, de l'étude pilote du SMSC (par le biais dune présentation à la fois thématique et chronologique) ;
- analyser les données recueillies dans ce cadre de l'étude pilote ;
- sur la base des résultats épidémiologiques et de l'analyse des difficultés rencontrées dans le cadre de la conduite de l'étude, discuter de la pertinence et la faisabilité du déploiement du SMSC au niveau national ;
- enfin, proposer des pistes de travail et émettre des recommandations en vue de la mise en place progressive du SMSC final.

I – HISTORIQUE, BASES ET PRINCIPES DU SMSC

1 Historique du Système national de surveillance multisources des cancers (SMSC)

1.1 Le dispositif de surveillance des cancers en France

Le système de surveillance des cancers en France repose principalement sur les registres des cancers spécialisés (par organe) et généraux, qui assurent l'enregistrement exhaustif des nouveaux cas de cancers dans une zone géographique délimitée (en général le département). Regroupés au sein du réseau Francim, les registres fournissent des données d'incidence observée dans les zones couvertes, ainsi que des estimations d'incidence aux niveaux national et régional. Ces estimations sont produites par modélisation statistique à partir des données d'incidence observées et des données de mortalité du Centre d'épidémiologie sur les causes médicales de décès (CépiDC) de l'Inserm. Les données des registres permettent la réalisation d'études de survie.
Les principaux indicateurs épidémiologiques (incidence et mortalité) sont mis en ligne sur le site de l'InVS, ainsi que les projections pour l'année en cours depuis 2009. L'InVS constitue la tête de réseau et co-finance, avec l'Institut national du cancer depuis 2008, le réseau des registres dans le cadre d'un programme de travail pluri-annuel InVS – INCa – Francim – service de biostatistiques des Hospices civils de Lyon (ce dernier hébergeant et gérant la base de données commune des registres).

Un registre est défini au sens du Comité national des registres (CNR) comme étant « un recueil continu et exhaustif de données nominatives intéressant un ou plusieurs événements de santé dans une population géographiquement définie, à des fins de recherche et de santé publique, par une équipe ayant les compétences appropriées».

Figure 1 - Carte présentant la situation des registres du cancer en France métropolitaine qualifiés par le CNR au 1er janvier 2010 (source Francim)
Ne figurent pas sur la carte : le registre national de surveillance du mésothéliome pleural, les deux registres nationaux de l'enfant ainsi que le registre de la Martinique.

Pour le recueil des cas de cancer, les registres se basent sur différentes sources de signalement. Les principales sources sont :

- les dossiers cliniques des établissements où le patient est traité ;
- les laboratoires d'anatomie et de cytologie pathologiques (ACP), où le diagnostic est posé et qui transmettent des comptes-rendus d'examen ;
- les Départements de l'information médicale (DIM) des établissements de soins publics et privés, qui transmettent la liste nominative des séjours par cancer de leur établissement ;
- les Caisses d'assurance maladie qui transmettent les noms des personnes mises en affection longue durée (ALD) pour cancer.

Ces sources sont complétées par la recherche des cas soignés hors zone de couverture du registre, notamment par le biais de la base PMSI nationale.

Pour assurer leur mission, les registres sont donc entièrement dépendants des sources qui leur signalent les cas, puis de l'accès aux dossiers médicaux dans les services de soins et en médecine libérale.

1.2 Les limites du dispositif face aux nouveaux enjeux de la surveillance

Si les registres constituent la référence en matière d'incidence des cancers, ils présentent néanmoins un certain nombre de limites : délai de n+4 années pour la mise à disposition des données[1] , couverture limitée à 20 % du territoire (sauf pour les registres nationaux de l'enfant - le Registre national des hémopathies malignes de l'enfant et le Registre national des tumeurs solides de l'enfant - qui couvrent 100 % de la population des moins de 15 ans), pas de données infra-régionales (donc pas de zonage géographique fin possible), coût très élevé (6 M€ en 2010).
Par ailleurs, les attentes des pouvoirs publics autour des nouveaux enjeux de la surveillance justifient à présent la mise en place d'un système au niveau national pour un certain nombre de localisations ; ces attentes se concentrent autour du suivi des programmes de dépistage organisé (sein, colon-rectum, col de l'utérus), du suivi des mesures de prévention de certains facteurs de risque infectieux (hépatites, papilloma virus humain - HPV), des expositions professionnelles.

1.3 Les travaux sur l'opportunité de la mise en place d'un dispositif de surveillance épidémiologique national des cancers

La réflexion sur le renforcement du dispositif de surveillance des cancers en France est issue des travaux réalisés par la commission multidisciplinaire mise en place et animée par l'InVS en 2000 dans le cadre de la mission d'évaluation des conséquences sanitaires en France de l'accident de Tchernobyl.
Un premier rapport publié par l'InVS en décembre 2001 [5-7] soulignait les limites du dispositif reposant sur les registres des cancers et constatait l'insuffisance de la couverture nationale pour l'enregistrement des cas. En effet, les registres des cancers ne couvraient alors que 13 % de la population française.
Dans un second rapport publié par l'InVS en 2003 [4], la commission analysait les différents dispositifs de surveillance nationale envisageables pour la population générale : registre national, système fondé sur la collecte centralisée des comptes-rendus ACP, déclaration obligatoire (DO) des cancers thyroïdiens et système à deux niveaux (routine et alerte).
Parmi ces quatre systèmes, les trois premiers ont été exclus :
- un registre national était jugé trop coûteux au regard des bénéfices escomptés ; cependant, la commission recommandait le principe du registre pour la surveillance des cancers de l'enfant et de l'adolescent ;
- un système national fondé exclusivement sur l'ACP était pertinent, mais nécessitait d'importants aménagements concernant la collecte des données ;
- la DO du cancer de la thyroïde était discutable au regard du bon pronostic de ce cancer.

La commission proposait donc :
- dans un premier temps, de développer et tester un système de surveillance à deux niveaux : enregistrement continu et permanent des cas incidents de cancer, compte tenu de la disponibilité des bases PMSI nationales, couplé à la collecte et l'analyse détaillée des cas en cas d'alerte (cluster) dans une aire géographique localisée ;
- à moyen voire plus long terme, de renforcer ce système par l'utilisation des données ACP dès que la collecte des données standardisées serait possible à un niveau national et que le problème de l'existence d'un numéro d'identifiant santé unique par individu serait résolu. La mise en place d'un tel numéro constituait un pré-requis nécessaire au chaînage des données, permettant d'assurer le recensement et la vérification des cas incidents provenant de différentes sources d'information.
La Commission soulignait le fait qu'aucun système de veille ne serait performant sans une collecte standardisée nationale des données ACP. Elle rappelait enfin la nécessité d'une parfaite communication entre les différents acteurs de ce système et d'un engagement politique et financier des structures publiques pour que le dispositif soit pérenne.
Ces recommandations dépassaient le cadre strict des cancers thyroïdiens pour concerner plus largement la surveillance de toutes les localisations cancéreuses.
Les conclusions sur la nécessité de travailler sur une collecte standardisée des données ACP ont été reprises par la profession des pathologistes, qui sous la coordination du Pr Brigitte Franc et en lien avec les épidémiologistes, a débuté une réflexion sur la standardisation des comptes-rendus ACP. Cette réflexion a abouti à la production d'une fiche ACP avec peu d'items (extraite du compte-rendu). Cette fiche établie pour le cancer de la thyroïde comprenait l'ensemble des items nécessaires au médecin prescripteur et pour la surveillance épidémiologique. Cette fiche a été utilisée comme outil de recueil dans le cadre de l'enquête EDF-Francim en 2002. La fiche et son manuel d'utilisation, ont ainsi été testés et améliorés (réalisation du codage de 4 600 comptes-rendus à partir de la fiche ACP thyroïde).

[1] Ce délai doit être ramené à n+3 années à partir de 2011 dans le cadre du programme de travail pluriannuel InVS – INCa – Francim – service de biostatistiques des Hospices civils de Lyon.

1.4 La priorisation de la surveillance épidémiologique des cancers dans les politiques publiques : les plans cancers gouvernementaux

Deux plans cancer ont été lancés depuis 2003 par la présidence de la République.
Le premier plan cancer gouvernemental 2003-2007 a confié à l'InVS, dans le cadre de la première mesure, la mission de renforcer les registres, augmenter la surveillance en milieu urbain et améliorer la représentativité de la population. Il recommandait également le développement d'un système de surveillance national : le Système de surveillance multi sources des cancers (SMSC).
Dans le cadre du premier plan cancer, des registres des cancers ont été qualifiés en Gironde et dans la zone urbaine de Lille. Le taux de couverture de 20 % obtenu est identique à celui du SEER programme américain (Surveillance Epidemiology and End Results), équivalent de notre dispositif de registre.

Le plan cancer 2009-2013 réaffirme l'importance de bâtir un système national d'épidémiologie des cancers plus exhaustif, plus réactif, plus prospectif. Il recommande en particulier la poursuite de la mise en place du SMSC, en tenant compte des nouvelles orientations prises dans le cadre de la relance des systèmes d'information en santé autour du dossier médical personnel (DMP).

2 Les bases du SMSC (justification, principe, objectif)

2.1 Le SMSC : une alternative à un registre national et au développement de nouveaux registres

Depuis plusieurs années, l'InVS est régulièrement sollicité par les pouvoirs publics et les élus locaux pour la mise en place de nouveaux registres des cancers au niveau départemental, afin de répondre à des inquiétudes locales vis-à-vis de possibles effets sanitaires à proximité de sites industriels pollués, d'usines d'incinération d'ordures ménagères, etc.
La politique française en matière de surveillance du cancer est actuellement de renforcer les moyens des registres existants plutôt que d'en créer de nouveaux. Par ailleurs, le CNR, s'appuyant sur les travaux de la Commission thyroïde, a jugé que le rapport coût/efficacité d'un registre national était défavorable. Le développement d'un registre national du cancer coûterait 25 à 30 millions d'euros par an, avec environ 300 à 350 ETP, sa montée en charge ne pourrait se faire en moins de cinq à dix ans.
Les autres grands pays d'Europe, à l'exception de l'Angleterre, ne disposent pas non plus de registres nationaux du cancer. A titre de comparaison, le registre anglais de la Tamise qui couvre environ 10 millions de personnes (semblable à la population de l'Ile-de-France) emploie à lui seul 70 personnes.
Le SMSC représente une solution alternative à un registre des cancers par un enregistrement automatisé de données individuelles anonymisées

2.2 Principe

Le principe du SMSC est d'appliquer sur tout le territoire, dans le cadre d'un système entièrement automatisé et passif et sur une base anonyme, la démarche utilisée par les registres des cancers pour l'enregistrement des cas incidents de cancer. Cette démarche repose sur le croisement de plusieurs sources d'information, qui permettent d'attester avec certitude du diagnostic de cancer et de son caractère incident (nouveau cancer, par opposition à une récidive ou à une rechute). Toutefois, à la différence du schéma envisagé pour le SMSC, les registres travaillent au niveau local, avec recherche active des informations (retour au dossier) et sur une base nominative.
Le SMSC repose ainsi sur le croisement de trois sources de données :
1/ les données issues du PMSI ;
2/ les données de l'Assurance maladie relatives aux exonérations du ticket modérateur pour cancer, correspondant à l'affection longue durée n°30 ou ALD30 ;
3/ les données issues des comptes-rendus d'anatomie et cytologie pathologiques (ACP).

2.2.1 Le choix des sources

- **Justification de l'utilisation des sources PMSI et ALD**

Dans tous les pays développés, bien que recueillies pour des raisons administratives et budgétaires et non à visée de santé publique, les bases médico-administratives alimentent les systèmes de surveillance, parfois même de manière exclusive, comme au Québec [8], du fait de leur disponibilité sans coût supplémentaire.
Les deux bases médico-administratives du PMSI et du Sniiram (Système national inter régimes de l'Assurance maladie) constituent, avec les certificats médicaux de décès, les principales sources d'informations « médicalisées » nationales exhaustives ; elles permettent de plus un suivi dans le temps. Ces sources présentent des qualités (donnés individuelles médicalisées, structurées et codées de manière standardisée) mais également des limites qui nécessitent d'être discutées avant leur utilisation dans un objectif épidémiologique. Des travaux ont été menés notamment à l'InVS et par le réseau Francim pour évaluer la qualité de ces sources, en déterminer les avantages et les limites et valider leur utilisation pour la production d'indicateurs d'incidence dite « médico-administrative » [9-13].

Le PMSI

La base de données médico-administrative que constitue le PMSI permet de connaître l'activité des établissements de soins publics et privés à travers un enregistrement des séjours hospitaliers. Au niveau national, une base de données est constituée de façon anonyme par l'Agence technique de l'information sur l'hospitalisation (ATIH) : la base des résumés de sortie anonymisés (RSA). Cette base de données, à couverture nationale, comporte des informations sur les diagnostics et les motifs d'hospitalisation pour chacun des séjours enregistrés.

Les données incluses dans les RSA comprennent notamment :

- le code Finess de l'établissement ;
- l'âge du patient à l'entrée dans le service ;
- le sexe du patient ;
- les mois et année de sortie ;
- la durée de séjour ;
- le code géographique de résidence ;
- les modes d'entrée et de sortie (domicile, transfert, décès) ;
- la provenance (si mode d'entrée est transfert ou mutation) ;
- la destination (si mode de sortie est transfert ou mutation) ;
- le diagnostic principal d'hospitalisation et les diagnostics associés : codés depuis 1997 selon la 10e version de la classification internationale des maladies (CIM-10) ;
- les actes réalisés, codés selon la classification commune des actes médicaux.

Dans le cadre de ses missions, l'InVS a reçu le 22 août 2002 une autorisation de la Commission nationale de l'informatique et des libertés (Cnil) d'accès permanent et de traitement des données issues de la base nationale du PMSI (demande d'autorisation n°902167). L'Agence technique de l'information hospitalière (ATIH) transmet chaque année au service des systèmes d'information (SSI) de l'InVS une copie intégrale de la base nationale des RSA[2] comprenant un identifiant anonyme généré par une procédure d'anonymisation appelée Foin (fonction d'occultation d'informations nominatives).

La procédure Foin est détaillée ci-après notamment au point 2.2.2 car elle est à la base de la constitution du numéro anonyme commun aux sources de données alimentant le SMSC.

Le Sniiram

Le Sniiram[3] est le fichier de consommations de soins (rapportées au bénéficiaire lui-même) de l'Assurance maladie. Il concentre, au niveau national, la quasi-totalité des informations issues de la liquidation des soins dans un entrepôt de données (datawarehouse) stockée, et en principe disponible, pour l'année en cours et les deux années précédentes (deux années glissantes). Au bout de deux ans, les données sont supprimées.

Pour les patients souffrant d'une ALD, le code en CIM10 de la pathologie qui a entraîné l'exonération du ticket modérateur est mentionné. En 2010, le Sniiram comprend les ALD du régime général (CnamTS), des sections locales mutualistes et du régime social des indépendants (RSI). En revanche, les ALD de la mutualité sociale agricole n'y figurent pas.

À terme, il est prévu que les données du PMSI hospitalier puissent être chaînées avec les soins dispensés en ville, ce qui devrait permettre de reconstituer des trajectoires entre l'hôpital et la ville.

L'utilisation du Sniiram à des fins de surveillance en santé publique par l'InVS n'est prévue pour l'instant que sous forme de données agrégées ce qui en limite fortement la portée et l'intérêt. Ainsi, l'InVS n'est pas autorisé à accéder en routine aux données individuelles, qui ne sont conservées que deux ans « glissant » dans le Sniiram et sont donc perdues ultérieurement, interdisant des analyses temporelles sur de longues périodes. En revanche, l'InVS a accès aux données individuelles de l'Echantillon généraliste de bénéficiaires au 1/97, mais cet accès ne permet pas d'étudier des pathologies peu fréquentes, et n'est pas représentatif au niveau régional ou départemental, ce qui ne permet pas d'analyse au niveau régional et infra régional, besoin rendu de plus en plus nécessaire, notamment avec la création des Agences régionales de santé (ARS). Chaque projet d'analyse impliquant une extraction des données individuelles du Sniiram doit ainsi faire l'objet d'une demande spécifique à la Cnil puis d'une extraction par la CnamTS, ce qui induit, dans les conditions de travail actuelles de ces deux institutions, des délais parfois incompatibles avec l'exercice des missions de l'InVS en situation d'alerte ou de saisine de la part de ses ministères de tutelle. Une demande d'autorisation de l'InVS à accéder aux données individuelles du Sniiram est en cours.

[2] La base comprend les RAS du PMSI-MCO (séjours en services de médecine, chirurgie, obstétrique) depuis 1998, du PMSI-SSR (séjours en soins de suite et de réadaptation [moyen séjour]) depuis 2005 et du PMSI-PSY (séjours psychiatriques) depuis 2007.

[3] Créé par la loi de financement de la sécurité sociale pour 1999.

- **La troisième source à construire : les données d'anatomie et cytologie pathologiques (ACP)**

L'anatomie pathologique est une spécialité médicale axée sur le diagnostic des lésions à partir de leur aspect morphologique. Les médecins pathologistes posent un diagnostic et un pronostic de maladie à partir de l'analyse sur lame des préparations de prélèvements tissulaires ou liquidiens provenant de l'activité chirurgicale, gynécologique, médicale ou radiologique. Cette discipline s'appuie sur des bases fondamentales d'anatomie normale, d'histologie et de cytologie pour identifier des anomalies morphologiques macroscopiques et microscopiques et sur des techniques d'histochimie, d'immunologie, de cytogénétique et de biologie moléculaire pour identifier des anomalies moléculaires dans les cellules ou les tissus. Les résultats sont transmis sous forme de compte-rendu textuel au médecin prescripteur.
L'ACP permet d'attester du diagnostic de cancer dans la grande majorité des cas (sauf tumeurs inopérables et hémopathies malignes) et constitue donc une source de données médicales fiable, indispensable pour la surveillance épidémiologique et au-delà, pour tout le champ de la cancérologie (hors hématologie).

Le plan cancer 2003-2007 reconnaît cette place indispensable des données ACP, en particulier dans la stratégie de prise en charge des patients et d'amélioration de la qualité des soins, qui fait l'objet de plusieurs projets du plan :

- le projet de dossier communiquant cancer (DCC), outil de communication des réseaux régionaux de cancérologie, en particulier pour la préparation des réunions de concertation pluridisciplinaires (RCP) ; ce projet, piloté initialement par l'INCa, prévoit l'alimentation du DCC par les différents comptes-rendus d'examens du patient, et notamment le compte-rendu ACP ;
- le projet de standardisation des comptes-rendus (CRFS), également piloté par l'INCa.

Dans le cadre du SMSC, les sources médico-administratives du PMSI et du Sniiram doivent être complétées par la source ACP. En effet, aucune des trois sources ne suffit à elle seule pour attester à la fois du diagnostic et du caractère incident de tous les cas de cancer ; seul le croisement des trois permet d'éliminer un maximum de cas prévalents, à condition de disposer de données historiques pour chacune d'elles.
Toutefois, contrairement aux bases PMSI et Sniiram, il n'existe **ni centralisation, ni standardisation ni structuration des données ACP** ; cette source est donc **à construire *de novo*** dans le cadre du projet.

2.2.2 La clé d'appariement : le numéro anonyme Foin

Le croisement des trois sources nécessite l'utilisation d'un identifiant anonyme unique par patient qui soit commun aux trois bases. Les bases nationales PMSI et Sniiram disposent, pour leur chaînage, d'un numéro d'anonymat commun de 17 caractères, non signifiant, généré par la Fonction d'occultation d'informations nominatives (Foin), propriété de la Caisse nationale d'assurance maladie des travailleurs salariés (CnamTS).
La procédure Foin est une librairie d'anonymisation qui comprend un algorithme informatique associé à un code ou secret d'anonymisation. Ce secret est paramétrable, de sorte que la procédure peut être cédée par la CnamTS à des partenaires extérieurs afin de pouvoir être mise en œuvre dans le cadre de projets indépendants.
Pour renforcer l'anonymisation, la procédure est passée deux fois sur le même identifiant, avec un secret différent à chaque niveau (émetteur et récepteur), le secret étant détenu par une autorité différente et indépendante ; il s'agit du principe de double anonymisation irréversible.
La fonction d'anonymisation présente les caractéristiques suivantes :

- l'irréversibilité ;
- l'agrégation, permettant un suivi dans le temps et dans l'espace d'un individu (à un identifiant nominatif correspond toujours le même identifiant anonyme) ;
- un faible taux de collision (faible probabilité d'obtenir un même anonymat à partir de données nominatives différentes) ;
- un bon effet avalanche (deux données approchantes ont des anonymats très différents) ;
- ainsi que le caractère paramétrable, déjà évoqué ci-dessus, et qui permet, pour deux identifiants nominatifs identiques dans deux projets différents d'obtenir en sortie deux anonymats différents.

Dans le cas de l'anonymisation des bases PMSI et Sniiram, le premier secret détenu par l'organisme émetteur (respectivement l'établissement de santé et l'échelon local de l'assurance maladie) permet de générer un premier numéro d'anonymat (anoSNIIR1) au départ des données. Le second secret, détenu par l'organisme récepteur (respectivement l'ATIH et le Centre national de traitement de l'information de la CnamTS – le Centi) permet, à la réception des données, de générer l'anonymat définitif (anoSNIIR2) qui sera conservé au niveau des bases nationales PMSI et Sniiram.
Pour la génération de l'anoSNIIR 1, la librairie Foin est alimentée par trois variables : **le numéro de sécurité sociale (NIR[4]) de l'assuré, le sexe et la date de naissance du bénéficiaire**.
Dans le cadre du SMSC, il a été choisi d'**appliquer la procédure Foin aux données ACP** pour générer le même numéro d'anonymat que celui commun au PMSI et au Sniiram.

[4] NIR : numéro d'inscription au répertoire (d'identification des personnes physiques).

2.3 Objectifs initiaux du SMSC

Les objectifs du SMSC définis initialement, avant réalisation de l'étude pilote, étaient ambitieux.
L'objectif principal était de renforcer la surveillance épidémiologique des cancers, en permettant :

- de disposer de données d'incidence des cancers, avec une réactivité plus grande que les registres, pour tout le territoire et à une échelle géographique fine ;
- de suivre leur évolution dans le temps (observation des tendances) ;
- dans un but d'orientation et d'évaluation des politiques de santé (dépistage, prévention, réponse à des questions environnementales).

Le SMSC comptait également deux objectifs secondaires :

- pouvoir appuyer les investigations en cas d'alerte sanitaire (détection ou signalement d'un agrégat spatio-temporel de cas de cancer), en fournissant :
    - o des données de référence sur l'incidence des cancers à un niveau géographique relativement fin (commune), afin de valider ou non l'excès de cas signalé ;
    - o des éléments sur d'éventuels autres cas non signalés dans l'aire géographique concernée et pouvant être rattachés à l'agrégat.

  Cet objectif d'alerte justifiait :
    - o éventuellement le retour au cas (en cas d'identification dans la base d'autres cas de cancer – même type histologique, même commune de résidence, même tranche d'âge, même période de diagnostic…) afin de mener les investigations appropriées et prendre le cas échéant des mesures de protection des populations exposées. Cette situation, bien qu'exceptionnelle, relève du champ de compétence de l'InVS et entre dans le cadre de ses missions ;
    - o la conservation illimitée des données du SMSC.
- fournir une aide à la conduite d'études épidémiologiques et de projets de recherche.
  Cet objectif répondait à une demande forte des départements scientifiques de l'InVS et des équipes de recherche partenaires de pouvoir alimenter les bases de données de cohortes avec les données du SMSC.

II – PREPARATION DE L'ETUDE PILOTE DU SMSC

1 Les objectifs de l'étude pilote

1.1 Objectifs généraux

L'étude pilote avait pour objectif d'étudier la pertinence et la faisabilité de la mise en place du SMSC pour une première localisation : le cancer de la thyroïde.

1.2 Objectifs opérationnels

Les objectifs opérationnels étaient les suivants :

- monter un réseau multi-partenarial avec les acteurs concernés par le projet ;
- mettre en place une solution informatique complète pour la remontée des données ACP. Cette solution devait comprendre le recueil, l'anonymisation, la transmission sécurisée des données ACP (à partir d'un nombre restreint de structures ACP et sur une période de temps donnée), jusqu'à l'InVS puis l'appariement avec les données PMSI et ALD30, sur la base du numéro d'anonymat commun, pour constituer une base de données test ;
- récupérer les données du PMSI et du Sniiram avec le numéro d'anonymat commun ;
- évaluer la qualité de l'appariement des trois sources ;
- tester l'objectif d'alerte ;
- identifier les risques au déploiement ;
- proposer les modalités les plus pertinentes pour le système en configuration nationale.

1.3 Ce que le pilote n'avait pas vocation à tester

Il est important de préciser que le pilote n'avait pas d'objectif épidémiologique. L'absence de registre en activité dans les régions choisies au moment du lancement de l'étude (voir chapitre matériel et méthodes), c'est-à-dire de gold-standard, ne permettait pas d'envisager de tester l'exhaustivité du système. Par ailleurs, compte tenu du caractère limité du recueil des données ACP (huit structures pilotes pour les deux régions) et de l'absence de données historiques pour les trois sources (inexistantes pour la source ACP, non demandées pour les sources PMSI et ALD30), la réflexion sur la sélection des cas incidents au sein de la base SMSC croisée n'était pas non plus un objectif de l'étude pilote.
Enfin, un des deux objectifs spécifiques du système, à savoir l'alimentation les bases de données de cohortes avec les données du SMSC n'a pas été testé (recueil trop limité des données ACP).

2 Périmètre de l'étude pilote

L'étude pilote a débuté fin 2004 avec le recrutement, à partir du 24/11/2004, d'un médecin épidémiologiste temps plein en charge de la coordination de l'étude et la tenue du premier comité de pilotage le 25/11/2004.

2.1 Les régions

L'étude ciblait spécifiquement deux régions pilotes, l'Ile-de-France et le Nord-Pas-de-Calais, définies par le plan cancer au regard 1/ des priorités de santé publique dans ces deux régions et 2/ du développement, en parallèle, d'études de faisabilité pour la mise en place de registres de cancers en zones urbaines dans ces régions – le SMSC devant constituer une alternative au registre en cas d'échec.
La population était de 11 291 000 habitants pour l'Ile-de-France et 4 028 000 habitants pour le Nord-Pas-de-Calais (estimations recensement 2004), ce qui représentait en 2005 respectivement 19 % et 7 % de la population française.

2.2 Le choix du cancer thyroïdien

Le choix de cette localisation faisait suite aux travaux de la Commission Tchernobyl. En raison de la faible fréquence des cancers thyroïdiens (9 000 nouveaux cas/an avec une incidence[5] de 5,4 pour 100 000 chez l'homme et 16,3 pour 100 000 chez la femme – projections pour l'année 2010 disponibles en ligne sur le site de l'InVS : http://www.invs.sante.fr), associée à leur faible létalité (380 décès/an soit une mortalité de 0,2 pour 100 000 chez l'homme et chez la femme – projections 2010), ils ne constituaient pas un problème prioritaire de santé publique. Cependant, leurs caractéristiques en faisaient un cancer pertinent pour tester ce nouveau modèle de surveillance nationale : faible fréquence donc faible volume d'enregistrement, diagnostic anatomo-pathologique et traitement chirurgical quasi systématiques.

[5] Taux standardisés monde pour 100 000 personnes-années.

Pour les deux régions pilotes, l'incidence du cancer de la thyroïde, en termes de nombre de nouveaux cas, était estimée à :

- 208 pour la région Nord-Pas-de-Calais en 2001 (estimation de l'Assurance maladie : régime général, mines, MSA, professions libérales) ;
- 750 pour la région Ile-de-France en 2000 (chiffres issus de la présentation d'Hubert Isnard au Copil SMSC de novembre 2004).

2.3 Partenaires et circuit des données

Un certain nombre de partenaires ont été sollicités pour participer à l'étude pilote : des structures ACP (au niveau local), des partenaires régionaux (le Centre de regroupement en informatique et statistique en ACP de la réunion Ile-de-France et les Directions régionales des affaires sanitaires et sociales d'Ile-de-France et du Nord-Pas-de-Calais) ainsi que des partenaires nationaux : la CnamTS et l'Institut national de la santé et de la recherche médicale (Inserm).

2.3.1 Les structures ACP pilotes

- **Evaluation de l'activité en cancérologie thyroïdienne et de l'état du parc informatique des structures ACP**

Dans certaines régions, les pathologistes sont organisés en associations, les Centres de regroupement informatique et statistique en ACP (Crisap), qui ont pour mission la mise en commun et l'analyse de données ACP au niveau régional. Pour les pathologistes, la participation à un Crisap est volontaire et les données recueillies par les Crisap ne concernent pas la totalité des examens ACP. Les Crisap ne fonctionnement pas dans l'ensemble des régions. Dans les régions couvertes par l'étude pilote, il existait un Crisap en Ile-de-France (Crisapif) mais pas dans le Nord-Pas-de-Calais.
Compte tenu de l'existence du Crisapif et de la volonté de la fédération des Crisap (FCcrisap) de participer au projet, il a décidé de constituer un partenariat entre l'InVS et le Crisapif afin de :

- faire participer le Crisapif à l'étude pilote SMSC en tant que partenaire régional dans la remontée des données (participation à l'assurance qualité des fiches au niveau régional) ;
- réaliser une enquête conjointe Crisapif/InVS en Ile-de-France pour le recensement des structures ACP, l'évaluation de leur activité en cancérologie thyroïdienne et de leur équipement informatique.

Concernant l'enquête de recensement des structures ACP en Ile-de-France, un questionnaire a été élaboré conjointement par l'InVS et le Crisapif début 2005. Les questionnaires ont été envoyés le 21 avril 2005. Un courrier d'accompagnement, préparé et cosigné par le Crisapif et l'InVS, permettait d'informer les ACP sur cette collaboration Crisapif – InVS. Deux modes de réponse étaient prévus : renvoi d'un exemplaire papier ou remplissage en ligne (version électronique du questionnaire).
Les résultats, présentés lors de la réunion du Copil du 22 septembre 2005, étaient les suivants : en Ile-de-France, le Crisapif recensait 113 structures ACP employant 489 pathologistes. Parmi ces 113 structures, 94 (83 %) ont répondu. Soixante et onze structures déclaraient une activité en pathologie thyroïdienne : il s'agissait pour moitié de structures publiques, plutôt des « grosses » structures employant 5 pathologistes en moyenne (de 1 à 20) et avec une activité moyenne à importante (+ de 5 000 actes/an) pour 82 % d'entre elles.
On notait une activité en cancérologie thyroïdienne plutôt « dispersée » : 32 % des structures diagnostiquaient moins de 5 cancers thyroïdiens par an, 40 % entre 5 et 20 et 23 %, de 20 à plus de 50.
Les structures présentaient un bon niveau d'équipement informatique, avec l'existence d'un système de gestion de laboratoire pour 92 % d'entre elles, distribué par l'un des quatre principaux éditeurs de logiciels du marché dans 86 % des cas (dont Infologic pour 47 % d'entre elles). Elles disposaient d'un accès Internet dans 95 % des cas, dont du haut débit pour 89 % d'entre elles.
Dans le Nord-Pas-de-Calais, en l'absence de Crisap actif et compte tenu du petit nombre de structures ACP, un état des lieux a été réalisé sur la base des données dont disposait la Drass à partir du fichier Adeli 2004. Dans la région, 25 lieux d'exercice ACP (6 dans le Pas-de-Calais, 19 dans le Nord) étaient recensés dont les deux tiers en secteur privé, employant 67 pathologistes. La métropole lilloise regroupait 44 % des lieux d'exercice. Ces structures employaient 76 médecins pathologistes (9 dans le Pas-de-Calais, 67 dans le Nord dont 28 pour le seul CHR). Le niveau d'informatisation des structures était présenté comme très bon, avec une forte représentation de la société Infologic.

Tableau 1 – Editeurs de logiciels ACP – Représentation au niveau national en 2005, étude pilote SMSC, 2004-2010

Nom de la société	Nombre de sites ACP équipés
Infologic	200
Technidata	120
Log-system	60
MB Concept	30

Cabinet Richard (anciennement SP-Soft)	30
Total	**440**

- **Sélection des structures ACP dans les deux régions pilotes**

Afin de réaliser l'étude pilote dans les conditions les plus proches possibles de la situation des ACP au niveau national, le panel des structures volontaires devait présenter les caractéristiques suivantes :

- activité en cancérologie thyroïdienne ;
- représentativité des différents secteurs d'activité (non au sens statistique du terme, mais présence dans l'échantillon d'au moins une structure par région exerçant dans les secteurs d'activité suivants : public, libéral, privé participant au service public hospitalier type centre de lutte contre le cancer (CLCC)) ;
- représentativité en termes de taille de la structure (grosse structure / petite structure) ;
- équipement informatique approprié.

Sur cette base, et en tenant compte des résultats de l'enquête conjointe Crisapif – InVS, les structures en IdF ont été sélectionnées sur proposition du Crisapif puis après contact et accord des différents sites pressentis.
Dans le NPdC, une réunion de présentation de l'équipe et du projet a été organisée à la Drass du Nord-Pas-de-Calais le 6 juin 2005. Sur les 12 pathologistes présents, exerçant en secteurs public, libéral et public PSPH, 3 se sont portés volontaires pour participer à l'étude pilote.
Au total, pour les deux régions pilotes, 8 structures ACP volontaires ont été sélectionnées.

Tableau 2 – Structures ACP pilotes et secteur d'activité, étude pilote SMSC, 2004-2010

	Structure	Secteur d'activité
Ile-de-France	Institut Gustave Roussy	Privé PSPH (centre de lutte contre le cancer)
	GH Pitié-Salpêtrière	Public
	Hôpital Ambroise Paré	Public
	Cabinet de pathologie Tolbiac	Libéral
	Institut de pathologie de Paris	Libéral
Nord-Pas-de-Calais	CHRU de Lille	Public
	GH de l'Institut catholique de Lille	Privé PSPH
	Laboratoire de Valenciennes	Libéral

2.3.2 Les autres partenaires

La remontée des données ACP était prévue selon trois niveaux :

- un niveau local (avec les structures ACP pilotes, en charge du recueil des données) ;
- un niveau régional, dont la composition et le rôle étaient à déterminer (intégrant le Crisapif en Ile-de-France) ;
- un niveau national avec, *in fine*, l'InVS en charge de la réception et de l'analyse des données.

L'ensemble des étapes, depuis le recueil des données par les ACP jusqu'à l'alimentation de la base SMSC au niveau national devait pouvoir être informatisé et automatisé. **Le circuit détaillé avec toutes les étapes d'anonymisation et de traitement des données est présenté au paragraphe « 3. Définition du schéma d'architecture d'anonymisation ».**

Le niveau régional

Le niveau régional était composé :

- en Ile-de-France, du Crisapif et de la Drass Ile-de-France (Drass 75) ;
- dans le Nord-Pas-de-Calais, en l'absence de Crisap, de la Drass 59 qui jouait les deux rôles (Crisap et Drass) ; elle était équipée pour ce faire de deux serveurs informatiques distincts situés dans des lieux géographiques différents.

Cette séparation du niveau régional en deux structures, chacune équipée d'un serveur informatique, permettait de recevoir les données ACP selon deux flux distincts :

- un flux correspondant aux données « médicales » du compte-rendu ACP (adressées sur le serveur hébergé par le Crisapif et un serveur Drass 59),
- un flux correspondant au numéro de fiche ACP anonymisé ou données « identifiantes » (adressé sur les serveurs Drass 75 et 59 dédiés).

Le rôle joué par le niveau régional a essentiellement consisté en l'anonymisation des données ACP et la centralisation des flux (la CnamTS souhaitant recevoir un flux de chaque région plutôt que de toutes les structures ACP).
Le rôle prévu initialement d'élaboration de la procédure d'assurance qualité et de la production et du suivi des indicateurs qualité des fiches n'a pas été mis en œuvre, compte tenu du rythme effectif de transmission des données pendant l'étude (envoi unique de toutes les données saisies), différent de celui envisagé initialement (envois itératifs à chaque nouveau diagnostic posé).

Le niveau national (hors InVS)

- **La CnamTS**

Le Centi de la CnamTS (chargé de la gestion du Sniiram) a participé à l'étude pour l'anonymisation des données identifiantes transmises par les Drass, puis leur retransmission vers l'InVS.

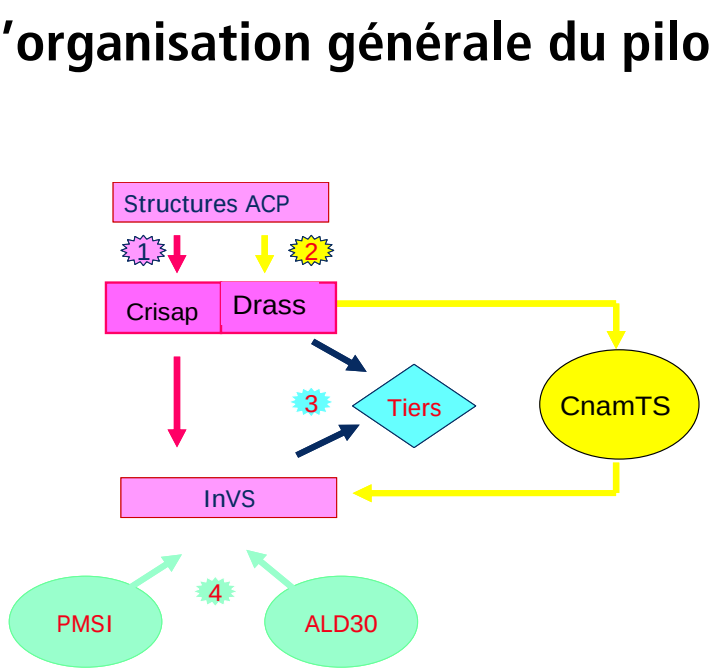
- **Le tiers de confiance : l'Institut fédératif de recherche (IFR) 69 de l'Inserm**

L'investigation de clusters (agrégats spatio-temporels de cas de cancer), qui constituent une situation d'alerte sanitaire, peut nécessiter exceptionnellement pour l'InVS de retourner au cas (par exemple pour rechercher l'adresse exacte ou la notion d'exposition).
Afin de concilier cette mission de l'InVS (figurant parmi les objectifs de l'étude), à savoir tester le retour au cas en cas d'alerte, avec les impératifs de sécurité de la CnamTS concernant l'utilisation de la procédure Foin (double anonymisation irréversible empêchant tout retour au patient), il est apparu indispensable d'introduire un tiers de confiance dans le schéma de circulation des données.

L'IFR 69 de l'Inserm a été sollicité pour jouer le rôle de tiers de confiance de l'étude. Il était chargé de conserver un fichier de correspondance entre « l'identifiant » du patient en clair (le numéro de fiche ACP) et le numéro d'anonymat final généré, pour ce patient, au niveau de la base croisée test.

Les relations entre l'InVS, le tiers de confiance et l'assurance maladie ont été contractualisées sur la base d'une convention tripartite, précisant les modalités de conservation du fichier de correspondance et les conditions d'accès au fichier pour l'InVS.

Figure 2– Partenaires et flux de données, étude pilote SMSC, 2004-2010



2.4 Sources de données

2.4.1 Source ACP

- **Définition de cas**

La définition de cas était basée sur le diagnostic de cancer de la thyroïde posé par l'ACP, qui constituait la source princeps.
Les cas correspondaient aux patients résidant dans les deux régions pilotes et pour lesquels un cancer de la thyroïde était diagnostiqué au sein de l'une des huit structures ACP. Etaient exclus les tumeurs bénignes de la thyroïde, les tumeurs border-line, les diagnostics de second avis (relecture de lame par un expert pour confirmer ou infirmer un diagnostic posé par un autre pathologiste), ainsi que les diagnostics sur pièce opératoire concernant des patients hospitalisés en dehors de l'établissement (dans le cas des structures ACP publiques).

- **Volets de l'étude**

L'étude s'est décomposée en deux volets distincts :

- **un volet rétrospectif** avec recueil de tous les cas de cancer thyroïdien (patients domiciliés) ayant fait l'objet d'un diagnostic ACP entre le 1er janvier et le 31 décembre 2004 ;
- **un volet prospectif** avec recueil de tous les cas de cancer thyroïdien (patients domiciliés) ayant fait l'objet d'un diagnostic ACP entre le 1er février et le 1er mai 2009.

À chacun des volets était attaché un objectif spécifique :

- volet rétrospectif : son objectif était de tester la génération du numéro d'anonymat commun pour les trois sources de données et de tester leur appariement au niveau national. Il était nécessaire de définir une année pour laquelle chacune des trois sources pouvait être disponible. Au moment du lancement de l'étude pilote, seule l'année 2004 était commune aux trois sources de données ;
- volet prospectif : son objectif était de tester 1/ l'utilisation de l'application informatique dans la pratique courante du pathologiste (déclaration prospective des cas de cancer diagnostiqués) et 2/ le fonctionnement en routine de la solution informatique d'anonymisation et de transfert sécurisé des données ACP entre tous les échelons du dispositif. Le volet prospectif n'avait aucun objectif d'appariement de données.

Le nombre de cas attendu estimé pour les huit sites pilotes était de l'ordre de 400 à 500 pour le volet rétrospectif, en appliquant les taux France entière en 2000 [2], et probablement supérieur à 125 pour le volet prospectif, compte tenu de l'augmentation continue de l'incidence des cancers de la thyroïde constatée en France comme dans tous les pays développés depuis les années 1970 [4].

- **Données recueillies**

Le support du recueil des données ACP était la fiche épidémiologique thyroïde (annexes 1 et 2), qui comprenait deux versions selon le secteur d'activité public ou privé de la structure ACP émettrice.

Pour chaque cas de cancer thyroïdien, les informations suivantes étaient recueillies.

Tableau 3 : Les informations de données ACP recueillies

	Variables
Identification de la fiche	Numéro FINESS de la structure ACP émettrice (pour les structures privées, un code était créé à partir du code Insee de la commune où la structure était implantée) Numéro de fiche (propre à chaque structure) Date de saisie de la fiche Nom du médecin déclarant
Données relatives au patient	Sexe du patient Date de naissance du patient Numéro patient : pour les structures ACP libérales : numéro de sécurité sociale de l'ouvrant droit et sa clé ; pour les structures ACP publiques : numéro d'identifiant patient dans l'établissement (NIP) et/ou numéro d'anonymat de premier niveau du PMSI (anoSNIIR1), à récupérer dans le système d'information hospitalier de l'établissement. Commune en clair et code postal de domicile
Données relatives au cancer thyroïdien	Date de l'enregistrement du prélèvement Nature de la pièce opératoire Poids et/ou taille de la pièce opératoire (lobe gauche, lobe droit) Nombre total de prélèvements sur la thyroïde examinée Nombre total de prélèvements sur la lésion Geste ganglionnaire ; si ganglions, préciser : nombre/envahissement si ganglions envahis, préciser : Ganglions lymphatiques régionaux homolatéraux Ganglions prétrachéaux / paratrachéaux / prélaryngés Autres chaînes cervicales ou médiastinales supérieures Taille de la lésion la plus grande Focalité de la lésion Latéralité de la lésion Capsule de la lésion ; si encapsulée, préciser : infiltration/dépassement Invasions vasculaires ; si présence, précisez si > 4 Invasions extra thyroïdiennes pTNM Numéro de l'édition du pTNM utilisée Diagnostic final (selon la codification CIMO -3) Second avis demandé ; si oui, préciser si diagnostic final posé après second avis Lésions associées (selon la codification CIMO -3) Diagnostic final du cancer associé (selon la codification CIMO -3) Commentaire libre

Parmi ces variables, certaines n'étaient pas transmises au niveau régional. Il s'agissait :

- du nom du médecin déclarant ;
- du numéro de sécurité sociale de l'ouvrant droit ainsi que sa clé (cas des structures ACP libérales) ;
- du commentaire libre.

Dans le cadre du volet rétrospectif (données 2004), chaque cas de cancer thyroïdien devait faire l'objet du remplissage d'une fiche épidémiologique ; les fiches, une fois renseignées, devaient être anonymisées avec la procédure Foin puis transmises à l'InVS en une seule fois *via* le niveau régional.
En cas d'information complémentaire non disponible au moment de la validation de la fiche (modification du diagnostic histologique suite à une reprise chirurgicale, totalisation d'exérèse avec curage ganglionnaire…), le pathologiste avait la possibilité de saisir une nouvelle fiche pour le patient concerné. Il était prévu que le niveau national procède à la gestion des doublons sur la base du numéro d'anonymat Foin et, dans le cas présent, ne conserve pour un même patient que la fiche portant la date de saisie la plus récente.

Pour le volet prospectif, le pathologiste devait remplir une fiche épidémiologique pour chaque nouveau cas de cancer thyroïdien diagnostiqué et transmettre la fiche validée de manière prospective dès validation. L'application informatique devait être paramétrée pour un envoi automatique des fiches validées de façon hebdomadaire (permettant au pathologiste de pouvoir y revenir pour les compléter ou les modifier en cas de besoin).
Chaque pathologiste devait conserver un fichier électronique « structure » comportant l'ensemble des numéros de fiches validées transmises, leur date d'envoi, et l'identité des patients correspondante, à des fins d'alerte et d'exercice du droit d'accès et de rectification des patients.

2.4.2 Source ALD30 (données du Sniiram)

La requête déposée auprès de la CnamTS, dans le cadre de la convention InVS/CnamTS, devait permettre d'extraire du Sniiram les données individuelles correspondant aux patients domiciliés en Ile-de-France ou dans le Nord-Pas-de-Calais et bénéficiant en 2004 d'une exonération du ticket modérateur au titre de l'ALD pour un cancer thyroïdien (code CIM10 : C73). Pour chaque patient identifié, les variables collectées sont décrites dans le tableau 4.

Tableau 4 : Les informations ALD recueillies

Rubrique	Variables
Organisme de prise en charge	« Grand régime »
Bénéficiaire	Mois de naissance Année de naissance Sexe Département de résidence Commune de résidence Bureau distributeur (code postal)
Informations médicalisées bénéficiaire	Numéro d'ALD Codes pathologie (pathologie principale et associées en CIM10) Date de mise en ALD Date de fin d'ALD Motif d'exonération du ticket modérateur
Numéro d'établissement du lieu de soins	Finess géographique

La cession à l'InVS des données individuelles ALD30 issues du Sniiram comportant le numéro d'anonymat de niveau 2 (Foin2) était conditionnée 1/ à l'accord de la Cnil et 2/ à la signature d'un contrat de cession de données. Les ALD30 devaient être transmises à l'InVS via la solution informatique mise en place dans le cadre de l'étude, car elles étaient transmises en format NTE (norme Camel de l'Assurance maladie) et nécessitait donc l'installation à l'InVS de Camel pour la lecture et le décryptage des données.

2.4.3 Source PMSI

Les données du PMSI étaient disponibles à l'InVS. Dans le cadre du volet rétrospectif de l'étude, ont été extraits de la base 2004 les RSA dont le code CIM10 correspondait à un cancer de la thyroïde (C73) en diagnostic principal, associé ou relié, pour les patients domiciliés en Ile-de-France (départements de résidence : 75, 77, 78, 91, 92, 93, 94, 95) ou dans le Nord-Pas-de-Calais (départements de résidence : 59, 62).
Pour chacun de ces RSA, les variables suivantes ont été extraites : l'âge, l'année de sortie (2004), le mois de sortie, le sexe, le code géographique de résidence, le numéro Finess de l'établissement de prise en charge, l'identifiant anonyme, le numéro de séquence, la version du RSA, le groupe homogène de malades (GHM), durée du séjour, les actes thérapeutiques, le type de diagnostic (principal, associé ou relié).

3 Définition du schéma d'architecture d'anonymisation

Compte tenu de la propriété de Foin, c'est le Centre d'étude sur la sécurité des systèmes d'information (Cessi) de la CnamTS qui a élaboré le schéma d'architecture d'anonymisation, basé sur la librairie Foin et le logiciel de sécurisation Camel[6], dans le cadre de l'un des 3 groupes de travail chargés de la réflexion préparatoire à l'étude pilote (voir II- 4, Méthodologie de conduite de l'étude).

3.1 Une solution distincte public/privé

Le premier schéma, qui prévoyait l'installation de la solution Foin chez tous les ACP quel que soit leur secteur d'activité, a été présenté à la Cnil dans le cadre d'une première demande de conseil en juillet 2005. Lors d'une réunion de travail entre l'InVS et la Cnil en septembre 2005, la Cnil a informé l'InVS de l'impossibilité pour les ACP exerçant en établissement de santé de récupérer le NIR du patient pour générer le numéro anoSNIIR1. En effet, la collecte et l'utilisation du NIR ne sont possibles qu'à des fins de facturation d'un acte, ce qui n'est pas le cas des pathologistes du secteur public ; hors ce champ, l'autorisation d'utilisation du NIR relève d'un décret en Conseil d'Etat. Etant donné les délais d'obtention de ce type d'autorisation, la Cnil a demandé à l'InVS d'envisager pour l'étude une solution d'anonymisation distincte public/privé en étudiant la faisabilité technique de récupérer, dans les systèmes d'information hospitaliers (SIH) des établissements participants, le numéro anonyme anoSNIIR1 (commun au PMSI et au Sniiram) et de l'appliquer aux données ACP.
La stabilisation du schéma d'architecture d'anonymisation a nécessité l'élaboration de plusieurs versions de travail entre avril 2005 et août 2006. Les dernières versions ont été produites par la CnamTS après l'autorisation de la Cnil (en juin 2006) pour :

- d'une part, répondre à la demande de la CnamTS de modifier le choix des partenaires régionaux et du tiers de confiance (qui initialement n'étaient ni les Drass, ni l'IFR 69 de l'Inserm) ;
- d'autre part, répondre à la demande de la Cnil que le tiers de confiance ne dispose pas du secret d'anonymisation de l'InVS.

Ces modifications, intervenant après l'obtention de l'autorisation, ont nécessité une information de la Cnil par le biais d'une nouvelle demande de conseil déposée en août 2008.

3.2 Le schéma retenu

Le schéma final reposait sur plusieurs objectifs et principes :

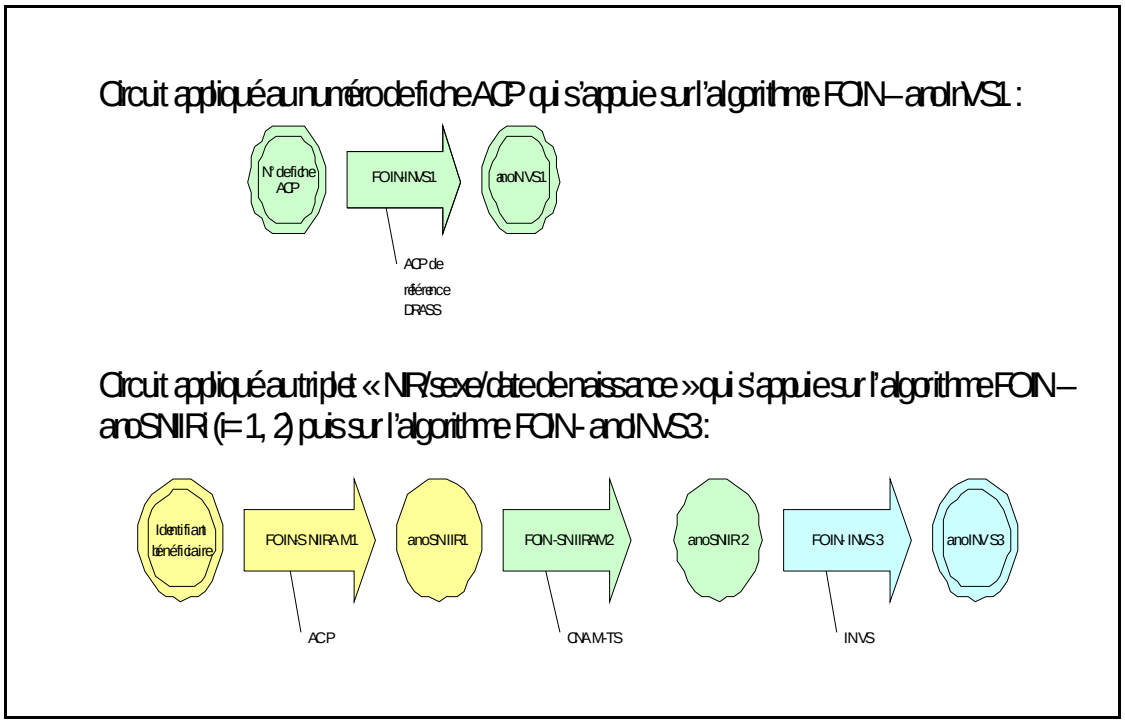
- pouvoir récupérer des informations « croisables » du Sniiram et du PMSI => extraction, à partir de ces deux entrepôts, des données nécessaires au projet ;
- ne pas compromettre la sécurité du Sniiram et du PMSI => ne pas diffuser le second niveau d'anonymisation du Sniiram et du PMSI ;
- garantir l'indépendance du projet InVS => utilisation de secrets spécifiques au projet ;
- répondre aux contraintes posées par la Cnil sur l'utilisation restreinte du NIR aux pathologistes du secteur privé.

L'InVS n'a donc pas eu la possibilité de générer lui-même le numéro anoSNIIR2 pour la source ACP. D'où la nécessité d'introduire le Centi de la CnamTS, gestionnaire du Sniiram, comme partenaire dans le circuit de remontée et d'anonymisation des données issues des structures ACP. Un secret spécifique a donc été attribué à l'InVS pour l'anonymisation finale des données de la base croisée test : le secret Foin3.
La procédure stabilisée prévoyait au final l'installation de Foin à tous les échelons du système avec deux circuits d'anonymisation en parallèle, chacun comportant deux niveaux de secrets :

- au niveau des structures ACP locales : secret de niveau 1 du Sniiram (anoSNIIR1 ou Foin1) ;
- au niveau régional (Crisap et Drass) : secret de niveau 1 propre à l'InVS (anoInVS1ou Foin InVS1) ;
- au niveau du Centi : secret de niveau 2 du Sniiram (anoSNIIR2 ou Foin2) ;
- au niveau de l'InVS : secret de niveau 2 propre à l'InVS (anoInVS3 ou Foin InVS3).

[6] Camel : chiffrement assurance maladie élargie.

Figure 3 – Circuits simplifiés d'anonymisation, étude pilote SMSC, 2004-2010



3.3 L'outil Camel

Dans le cadre de ses échanges de données anonymisées par la procédure Foin, la CnamTS utilise le logiciel de sécurisation des flux Camel. Ce logiciel, agréé par la Commission nationale de sécurité des systèmes d'information (CNSS), présente un certain nombre de fonctionnalités : chiffrement des données, signature électronique des envois permettant une authentification de l'émetteur, contrôle de l'intégrité et de la non-répudiation des données échangées. Camel permet ainsi d'engager la responsabilité des partenaires.
Il s'agit d'un outil robuste, utilisé également dans le secteur bancaire, mais contraignant pour plusieurs raisons :

- dans le champ sanitaire et médico-social, il est très peu utilisé et, le cas échéant, spécifiquement par l'Assurance maladie ;
- il est basé sur des technologies lourdes et relativement anciennes ;
- son utilisation implique :
    - o l'interfaçage obligatoire avec un logiciel qui gère l'automatisation des traitements (dont le choix sur le marché est très restreint) ;
    - o le formatage des données transitant par Camel selon la norme technique d'échange (NTE) élaborée par la CnamTS.

Dans la majorité des projets utilisant Foin (y compris hors champ de l'assurance maladie), la librairie Foin est encapsulée dans le logiciel Camel et ne peut être utilisée seule.

Pour la circulation des données ACP de l'étude pilote, la CnamTS a donc souhaité introduire Camel (intégrant Foin) dans le schéma d'architecture d'anonymisation.

3.4 La génération des secrets

La librairie Foin fonctionne avec des secrets d'anonymisation paramétrables, générés selon une procédure standardisée en présence de la CnamTS et des autorités administratives concernées. Ils doivent être conservés de manière sécurisée.
La réunion de génération des secrets Foin InVS1 et Foin InVS3 a eu lieu le 28 janvier 2008 en présence de représentants du Cessi de la CnamTS, de l'équipe projet InVS ainsi que des partenaires régionaux : Crisapif, Drass Ile-de-France et Drass Nord-Pas-de-Calais.

3.5 Détail des étapes d'anonymisation et de traitement des données (annexes 3 et 4)

3.5.1 Au niveau des structures ACP locales

Deux situations se présentaient selon le secteur d'activité des ACP :
1/ installation du module Foin pour générer l'anoSNIIR1 dans les structures libérales et des établissements privés PSPH ;
2/ mise en place de connecteurs informatiques entre la structure et le bureau des entrées de l'établissement pour la récupération de l'anoSNIIR1 généré par le logiciel MAGIC pour la génération des RSA (circuit d'anonymisation des données PMSI), dans le cas des structures ACP publiques.
Les données ACP devaient pouvoir être envoyées selon une périodicité hebdomadaire et suivre un double circuit :

- les données relatives à l'identification de la fiche, (le « numéro de fiche »), ainsi que l'anoSNIIR1 seraient adressées à la Cire (afin de permettre en fin de circuit l'appariement des données ACP, PMSI et Sniiram) ;
- les autres données de la fiche, hormis le NIR, qui ne circule à aucun moment dans le dispositif, seraient transmises à l'équipe ACP référente régionale pour être stockées dans une base temporaire, le temps d'effectuer le contrôle qualité et la validation, puis anonymisées avec Foin avant leur transfert au niveau national (envoi hebdomadaire). La base temporaire devait être détruite dès le transfert effectué de telle sorte qu'il ne persiste pas de base de données « ascendante » au niveau régional.

3.5.2 Au niveau régional

L'équipe ACP référente (Crisapif et Drass Nord-Pas-de-Calais) devait recevoir les données ACP couplées au numéro de fiche. Elle devait soumettre ces données à une procédure d'assurance qualité, les valider manuellement puis les anonymiser avec la fonction Foin InVS1 (génération du numéro anoInVS1).
Les Drass Ile-de-France et Nord-Pas-de-Calais devaient recevoir le numéro de fiche apparié au numéro anoSNIIR1. Elles devaient générer le numéro d'anonymat anoInVS1 à partir du numéro de fiche. Deux fichiers distincts devaient être préparés à l'issue :

- un premier fichier comprenant le numéro anoInVS1 apparié à l'anoSNIIR1 était adressé au Centi à Evreux (gestionnaire Sniiram) en vue du passage du niveau 2 sur l'anoSNIIR1 (transformation en anoSNIIR2) ;
- un second fichier comprenant le numéro anoInVS1 apparié au numéro de fiche était adressé au tiers de confiance pour l'alimentation du fichier de correspondance.

Au départ du niveau régional, trois circuits de transmission des données étaient donc mis en place :

- du Crisapif/Drass Nord-Pas-de-Calais vers l'InVS (anoInVS1 – données ACP) ;
- des Drass vers le Centi de la CnamTS (anoInVS1 – anoSNIIR1) ;
- des Drass vers le tiers de confiance (anoInVS1 – numéro de fiche).

3.5.3 Au niveau de la CnamTS

La CnamTS devait procéder à la génération de l'anoSNIIR2 à partir de l'anoSNIIR1 (anonymisation de deuxième niveau du Sniiram) ; l'anoSNIIR2 couplé à l'anonymat anoInVS1 était ensuite adressés à l'InVS.

3.5.4 Au niveau de l'InVS

- Concernant les données ACP : l'InVS disposait d'un premier fichier transmis par le Crisapif/Drass Nord-Pas-de-Calais comportant les données ACP couplées à l'anoInVS1, ainsi que d'un second fichier transmis par la CnamTS comportant l'anoInVS1 couplé à l'anoSNIIR2. Ces deux fichiers devaient être appariés grâce à l'anoInVS1, anonymisé ensuite par la fonction Foin InVS3 ce qui permettait de disposer en sortie des données ACP couplées à l'anoInVS3, constituant la base ACP nationale anonymisée.
- Concernant les données ALD30 et PMSI (année 2004 uniquement) : les données ALD et PMSI de test devaient être anonymisées selon la même procédure par la fonction Foin InVS3.

Les trois fichiers « données ACP – anoInVS3 », « données ALD30– anoInVS3 » et « données PMSI – anoInVS3 » pouvaient ensuite être appariées (pour l'année 2004) grâce à l'anoInVS3 pour constituer la base croisée test.

3.5.5 Au niveau du tiers de confiance

Le tiers devait disposer de deux fichiers : le premier, transmis par les Drass, comportait le couple « numéro de fiche - anoInVS1 ». Le second, transmis par l'InVS, comportait le couple « anoInVS1 - anoInVS3 ». Le tiers ne disposait d'aucun secret d'anonymisation ; en revanche, il devait disposer de Camel seul pour la réception des fichiers des deux fichiers. Son rôle était de pouvoir réaliser l'appariement des couples complémentaires sur la base du numéro commun anoInVS1 et conserver le fichier de correspondance comportant la liste des couples « numéro de fiche – anoInVS3 ».

3.6 La procédure de retour au cas dans le cadre de l'étude pilote

Afin de tester l'objectif d'alerte du projet SMSC, un exercice de retour au cas était prévu pendant l'étude pilote.

3.6.1 Cadre général de l'exercice

Il s'agissait d'un «test à blanc» de la solution informatique d'anonymisation et de transmission des données et non d'une situation réelle d'alerte sanitaire ; l'objectif de cet exercice était de s'assurer que l'ACP était capable, à partir du fichier de correspondance détenu par le tiers de confiance (couples numéro de fiche – anoInVS3), de retrouver un patient. En aucun cas il n'y avait de retour effectif à l'identité du patient ; il n'y avait donc pas de levée du secret professionnel dans le cadre de cet exercice.

3.6.2 Modalités de réalisation de l'exercice

Le retour au fichier de correspondance concernait exclusivement les données ACP 2004, transmises en une seule fois à l'InVS *via* les différents partenaires.

L'objectif était :

- d'évaluer la qualité de l'appariement des couples complémentaires (n° de fiche – anoInVS1) et (anoInVS1 – anoInVS3), attestant de la performance de la solution informatique ;
- de vérifier, à partir du fichier transmis par le tiers, que :
  - l'InVS pouvait identifier avec certitude le pathologiste déclarant (le numéro de fiche comprenait le code Finess de la structure émettrice) ;
  - le pathologiste ainsi identifié pouvait retrouver la fiche ACP transmise et, à partir de celle-ci, l'identité et les coordonnées du patient.

Dans le cadre de cet exercice, l'ACP ne devait pas communiquer ces données à l'InVS, mais devait préciser, pour chacun des cas, s'il avait pu ou non être identifié.

3.6.3 Échantillonnage et procédure de retour au cas

- Extraction par l'InVS de 100 cas par échantillonnage aléatoire à partir de la base des données ACP 2004.
- Transfert du fichier au tiers de confiance.
- Recherche par le tiers de confiance des couples numéro de fiche – anoInVS3 correspondants.
- Retour à l'InVS du fichier comportant le nombre de couples retrouvés.
- Identification par l'InVS de la ou des structure(s) ACP émettrice(s) à partir du numéro de fiche.
- Transfert à chaque structure ACP identifiée du fichier correspondant comportant le ou les numéro(s) de fiche(s).
- Identification par la ou les structure(s) ACP des cas correspondants au(x) numéro(s) de fiche(s).
- Transfert à l'InVS du fichier comportant, en regard de chaque numéro de fiche, la mention «le cas a pu être identifié» ou le cas échéant, «le cas n'a pas pu être retrouvé».

Cet exercice était prévu d'emblée et figurait dans le protocole initial de l'étude soumis à la Cnil. Toutefois, les modalités exactes n'en avaient pas été précisées. À la demande la CnamTS, ces modalités ont du être soumises à la Cnil dans le cadre d'une demande de conseil, déposée le 4 août 2008. L'accord a été obtenu au mois d'août 2009.

4 Méthodologie

4.1 Conduite de l'étude

La méthodologie de gestion de projet informatique a été appliquée à la réalisation de l'étude pilote SMSC, avec la mise en place de l'équipe projet, des instances de suivi du projet et la construction des partenariats.

4.1.1 Ressources

- **L'équipe projet InVS**

Le pilotage et la réalisation du projet ont été assurés par une équipe projet organisée en binôme « maîtrise d'ouvrage (MOA) – maîtrise d'œuvre (MOE) », avec au minimum un chef de projet MOA issu du Département des maladies chroniques et des traumatismes (DMCT) et un chef de projet MOE issu du Service des Systèmes d'information (SSI) de l'InVS.

L'équipe comptait au démarrage, côté MOA, un responsable du projet SMSC et un chef de projet – coordinateur de l'étude pilote. La MOA a été renforcée en avril 2008 par l'arrivée d'un chef de projet plus particulièrement en charge des travaux sur les données. Côté MOE, quatre chefs de projet se sont succédés entre fin 2004 et 2010. Si la composition de

l'équipe a varié au cours de la réalisation de l'étude pilote, elle a toujours compté au minimum deux épidémiologistes côté MOA représentant 1,3 équivalent temps plein (ETP).

- **L'assistance à maîtrise d'ouvrage**

Le projet a été identifié à deux reprises comme nécessitant une assistance à maîtrise d'ouvrage (AMOA) dans le cadre de la réalisation des projets prioritaires du schéma directeur des systèmes d'information (SDSI) de l'InVS, la première fois début 2005 (marché à bons de commande passé avec la société A), la seconde courant 2008 (marché similaire passé avec la société B).
La prestation AMOA de la société A a essentiellement consisté en la participation d'un médecin aux deux réunions du groupe de travail « système informatique » pour sa connaissance du PMSI.
La société B a été sollicitée en mai 2008 pour une AMOA sur le projet SMSC. Dans ce cadre, la société B a réalisé les prestations suivantes :
1/ Réalisation de neuf entretiens auprès de partenaires du projet (liste en annexe 5), sur la base d'une liste de noms fournie par l'équipe projet. Les grilles d'entretiens ont été élaborées avec l'équipe projet et ont été personnalisées en fonction des partenaires interrogés.
Ces entretiens ont été conduits entre octobre et décembre 2008 ; ils avaient pour objectifs de :

- mesurer, pour la personne interviewée, sa perception des enjeux et son intérêt pour le projet SMSC ;
- disposer d'éléments permettant d'enrichir le bilan de l'expérimentation (enjeux, réalisations, contraintes rencontrées…) ;
- mesurer la faisabilité de la phase de généralisation (enjeux, périmètre, contraintes identifiées, facteurs clés de succès…) ;
- identifier et obtenir un retour d'expériences de projets similaires de remontées d'informations ;
- poser (si nécessaire) des questions spécifiques à chaque structure interviewée.

Un tableau récapitulatif de la perception générale des partenaires interrogés au regard du projet figure en annexe 6.

2/ Préparation et animation d'un séminaire de réflexion le 12 février 2009 et installation, dans ses suites immédiates, de l'instance de gouvernance du projet (voir ci-dessous).

3/ Rédaction d'un rapport intermédiaire d'évaluation remis à l'InVS le 14 décembre 2009.

4.1.2 Comitologie

- **Le comité de pilotage (Copil)**

Le Copil a été constitué et réuni pour la première fois en novembre 2004. Impliquant un grand nombre de partenaires institutionnels mais également des partenaires intervenant directement dans le circuit de remontée des données ACP, il avait pour rôle de suivre les étapes de réalisation de l'étude et de se prononcer sur les choix de l'équipe projet InVS. Il s'est réuni au total huit fois entre 2004 et 2010.
Sa composition a peu varié dans le temps. A titre d'exemple, elle était la suivante lors du dernier Copil du 10 juin 2010 : CnamTS, INCa, Agence des systèmes d'information partagés de santé (ASIP Santé), Cnil, syndicat des médecins pathologistes français (SMPF), pathologistes des secteurs public et libéral, Crisapif, Inserm (U605), réseau Francim, conseil national de l'ordre des médecins, ATIH, Département d'information médicale général de l'Assistance publique-Hôpitaux de Paris…

- **La gouvernance (instances internes InVS)**

Devant les retards et les difficultés rencontrées lors de la réalisation de l'étude, un **comité stratégique** (CoStrat) a été mis en place en mars 2009, impliquant la direction générale (DG), la direction scientifique, les responsables de services ainsi qu'un représentant des départements scientifiques intéressés par le projet.
Les réunions du CoStrat se sont déroulées à un rythme mensuel dès leur démarrage ; elles étaient animées par le chef de projet – coordinateur de l'étude pilote. L'objectif de ces réunions était le suivi stratégique et la prise de décision concernant notamment :

- le suivi de la prestation pour la mise en place de la solution informatique de remontée des données ;

- les relations avec les partenaires institutionnels (l'INCa, l'Asip santé[7]) et la profession des pathologistes autour de l'avancée des autres projets du plan cancer basés sur les données ACP et pilotés par l'INCa et l'ASIP Santé : la mise en place du dossier communicant de cancérologie (DCC), service spécialisé du dossier médical personnel (DMP) et le projet de standardisation des comptes-rendus ACP (définition des items minimaux ACP devant figurer dans un compte-rendu).
Dans le cadre du schéma directeur des systèmes d'information (SDSI) de l'InVS, un **comité de gestion** (CoGEs) a été mis en place début 2006.

[7] L'Asip santé voit le jour en septembre 2009 dans les suites du GIP-DMP.

Son objectif était de suivre, dans le cadre de réunions mensuelles, l'avancement des projets stratégiques et complexes de l'Institut, depuis l'étape de consultation jusqu'à la mise en production des applications développées. Sa configuration a évolué au cours du temps ; dans sa forme actuelle, la composition du CoGes était la suivante : un membre de la DG, des responsables systèmes d'information (directeur du service et responsable du pôle applicatif), des responsables financiers (chef de service et responsable du pôle marché), ainsi que le chef de projet MOA de tous les projets SDSI.
Le projet SMSC a été présenté en CoGes dès début 2006. Un représentant du DMCT puis de la MOA a participé pour le SMSC à ces réunions à partir de 2007, et ce jusqu'au mois d'avril 2010, date de fin de la prestation informatique et de l'étude. Un retour d'expérience (RETEX) de l'étude pilote SMSC a été présenté le 1er juin 2010.

4.1.3 Groupes de travail multidisciplinaires

Débutée lors des travaux de la commission Tchernobyl, la construction des partenariats s'est échelonnée entre 2004, avec la constitution du Copil et la mise en place des différents groupes de travail et 2010, avec la signature d'une convention de partenariat entre l'InVS et l'ASIP Santé.
Afin de contribuer au protocole, trois groupes de travail (GT) pluridisciplinaires ont été mis en place début 2005 et ont permis la finalisation des principaux documents administratifs et techniques relatifs à l'étude ;

- un groupe « anonymisation », pour l'élaboration du schéma d'architecture d'anonymisation et des flux, en vue de la préparation du dossier d'autorisation à la Cnil ;
- un groupe « pôle régional », chargé 1/de réfléchir à la composition et aux modalités de fonctionnement du niveau régional et 2/de proposer les bases du cadre contractuel avec les partenaires (conventions, chartes et contrats) ;
- un groupe «système informatique», chargé de préparer le cahier des charges en vue du dossier de consultation pour la mise en place de la solution informatique de recueil, anonymisation et transfert sécurisé des données ACP vers l'InVS.

4.2 Méthodologie de sélection des données PMSI, de croisement et d'analyse des données médico administratives et anatomopathologiques recueillies

4.2.1 Algorithme de sélection des données du PMSI

La sélection des cas de cancer de la thyroïde dans la base PMSI 2004 était basée sur un algorithme combinant le code CIM10 correspond à un cancer thyroïdien (C73) en diagnostic principal, associé ou relié, et les codes des actes chirurgicaux du catalogue des actes médicaux (CdAM) et de la classification commune des actes médicaux (CCAM) associés pour les patients domiciliés en Ile-de-France ou dans le Nord-Pas-de-Calais.

4.2.2 Processus de croisement des bases

Le croisement des données des trois sources a été réalisé en plusieurs étapes :
- **Première étape :**
Les données de la base PMSI ont été croisées avec celles des ALD30 après suppression des doublons et des identifiants non valides (identifiants à caractère non exploitable). Ce croisement réalisé à partir des identifiants identiques a généré une base commune aux deux sources avec laquelle les données ACP sont croisées.
- Deuxième étape :
La base commune issue du croisement des données PMSI et ALD30 a été croisée avec les données ACP. Ce croisement a permis d'obtenir une base de données communes aux trois sources d'une part et une base de données ACP non retrouvées (données absentes) dans la base croisée PMSI/ALD30 d'autre part.
- Troisième étape (croisement des données absentes de la base commune ACP/PMSI/ALD) :
Les données ACP absentes de la base commune ACP/PMSI/ALD30 ont été croisées d'une part avec les données PMSI absentes du croisement PMSI/ALD30 et d'autre part avec les données ALD30 absentes du même croisement.

4.2.3 Constitution de la base d'analyse

La base finale d'analyse était composée des :

- données communes aux trois sources (n=191);
- données communes aux sources ACP/PMSI mais pas aux ALD (n=152) et ACP/ALD30 mais pas au PMSI (n=11) à partir des données non appariées entre les trois sources;
- données ACP non appariées ni aux données PMSI ni aux données ALD30 (n=69).

4.2.4 Analyse des données

Une analyse descriptive des variables de la base constituée a été effectuée.
Une recherche de cohérence entre les sources concernant le sexe, l'âge et le lieu de résidence a été réalisée.
Le taux d'appariement (Nombre de cas ACP retrouvés dans les autres bases) des observations croisées entre les sources a été calculé, d'abord entre les trois sources puis entre les sources PMSI/ACP et ALD30/ACP.
Le taux d'appariement des données de la région du Nord-Pas-de-Calais du pilote a été comparé avec celui du registre de l'année 2005 de l'agglomération de Lille dont les sources de signalement des cas proviennent des données du PMSI, des ALD30, des ACP et de radiothérapie. Cette comparaison a concerné les cas signalés au registre par une au moins des trois sources PMSI, ALD30 ou ACP.
Des analyses bivariées et multivariées ont été réalisées afin d'identifier les facteurs liés à l'absence d'un cas dans la base PMSI ou ALD30.
Enfin, une analyse descriptive des cas ACP présents dans la base ALD30 mais absents dans la base PMSI (n=11), ou présents dans la base PMSI et absents dans la base ALD30 (n=152) ou absents dans la base PMSI et ALD30 (n=69) a été réalisée.
Toutes les analyses ont été réalisées avec le logiciel SAS version 9.1

4.3 Exercice de retour au cas

L'InVS dans sa mission d'alerte sanitaire pourrait être amené à investiguer d'une manière exceptionnelle des cas groupés de cancer. Dans ce contexte, le protocole de l'étude a prévu de tester la capacité du dispositif à retourner au cas. Deux étapes complémentaires permettaient de retrouver les cas.
Pour procéder à ce test, un tirage aléatoire simple sans remise de 100 fiches a été réalisé sur les 423 fiches ACP exploitables adressées par les huit structures pilotes à l'InVS. Les identifiants anonymes des fiches tirées ont été transmis dans un premier temps au tiers de confiance (IFR69) en vue de la recherche des numéros de fiches correspondants puis dans un second temps aux huit structures ACP pour la recherche et l'identification des dossiers des patients correspondants dans leur base informatique de comptes rendus. Les résultats binaires (trouvé/non trouvé) de la recherche ont été ensuite transmis à l'InVS pour analyse.

4.4 Evaluation de l'étude pilote

Il était prévu d'emblée dans le protocole Cnil qu'une évaluation de l'étude pilote soit conduite et, en fonction des résultats, décider 1/de poursuivre le projet et généraliser le dispositif testé au niveau national, 2/de poursuivre le projet mais sous réserve d'adaptations avant généralisation, ou 3/d'arrêter le projet.

- Les critères d'évaluation

<u>Critères généraux</u>
1/ qualitatifs : acceptabilité des pathologistes, fonctionnement général du système informatique.
2/ quantitatifs :

- pourcentage d'échec d'appariement par le numéro d'anonymat entre les données ACP et les données PMSI et Sniiram pour l'année 2004 ;
- pourcentage de fiches incluses dans la base multi sources ;
- pour les données recueillies en prospectif, délai d'arrivée des fiches à l'InVS.

<u>Critères spécifiques</u>
1/ atteinte des objectifs.
2/ niveau d'utilité au regard du dispositif de surveillance existant.
3/ simplicité de la solution mise en œuvre.
4/ flexibilité de la solution mise en œuvre.
5/ qualité des données recueillies.
6/ acceptabilité / adhésion des partenaires au dispositif.
7/ reproductibilité de la solution testée aux 500 sites ACP sur le territoire, stabilité (en termes de choix techniques, de maintenance de la solution).
8/ réactivité.

III – REALISATION DE L'ETUDE PILOTE

Tableau 5 : Synthèse des principaux jalons

Année	Jalons
2004	1er Copil : novembre 2004
2005	Groupes de travail préparatoires : janvier à juin 2005 2e Copil : avril 2005 1ère demande de conseil Cnil : juillet 2005 3e Copil : septembre 2005 1re consultation : novembre 2005
2006	4e Copil : mars 2006 Dépôt du dossier d'autorisation : avril 2006 Accord Cnil : juin 2006 Préparation du 2e cahier des charges : septembre – décembre 2006
2007	2e consultation : février 2007 5e Copil : juin 2007 Notification au prestataire / Réunion de lancement de la prestation informatique : juillet 2007 Spécifications fonctionnelles et techniques de la solution informatique : novembre 2007
2008	Développements et installations : janvier – décembre 2008 6e Copil : avril 2008 2e demande de conseil Cnil : août 2008
2009	Premiers croisements de données : février 2009 Avis favorable de la Cnil à la 2e demande de conseil : août 2009 7e Copil : septembre 2009 Première communication des résultats (colloque sur la surveillance des cancers) : octobre 2009
2010	Fin de l'étude pilote (signature du PV de fin de VSR avec réserves) : février 2010 Retour d'expérience (Retex) : juin 2010 8e et dernier Copil : juin 2010 Rendu du rapport d'étude : T4

1 Les travaux administratifs

1.1 Le dossier d'autorisation Cnil

La préparation du dossier d'autorisation Cnil a débuté avec les travaux du GT « anonymisation ».
Plusieurs réunions avec la Cnil et deux demandes de conseil ont été nécessaires, en préparation puis en cours de réalisation de l'étude. L'autorisation de la Cnil pour la réalisation de l'étude pilote a été obtenue le 8 juin 2006 sous le numéro 906108, puis la demande d'extension de l'étude déposée par l'InVS le 11 août 2008 a bénéficié d'un avis favorable le 17 août 2009.
Les étapes de la constitution du dossier sont détaillées ci-dessous :

- préparation du dossier d'autorisation :
  - **dépôt d'une première demande de conseil** le 08/07/2005 ;
  - réunion de travail Cnil/InVS le 21/09/2005 : demande de la Cnil de mettre en place une solution d'anonymisation distincte chez les ACP pilotes exerçant en établissement public (pas d'installation de la procédure Foin dans ces structures) (voir paragraphe 3.1 « Une solution distincte public/privé ») ;
  - soumission d'une nouvelle version le 08/11/2005, avec réponse favorable de la Cnil le 31/01/2006 ;
- **dépôt du dossier de demande d'autorisation** le 13/04/2006 : autorisation obtenue le 8 juin 2006 sous le numéro 906108 ;
- extension de l'étude pilote :
  - réunion avec la Cnil le 20/12/2007, à la demande de l'équipe projet InVS, pour faire état :
    1/des difficultés rencontrées pour l'anonymisation des données ACP dans les établissements publics pilotes (notamment dans la mise en œuvre de la solution technique envisagée) ;

2/ du changement de partenaires par rapport au protocole initial (niveau régional et tiers de confiance) ;
3/ à la demande de la CnamTS, des modalités d'exercice de retour au cas prévu mais non décrit précisément dans le protocole initial ;
4/de la demande d'extension du périmètre de l'étude (recueil des données ALD tous cancers, France entière, à partir de l'année 2006 et suivantes en vue de leur croisement avec les bases PMSI correspondantes).
La Cnil propose à l'équipe projet InVS de solliciter l'avis de la Cnil sur ces différents points par le biais d'une demande de conseil ;

- **dépôt d'une deuxième demande de conseil** le 11/08/2008 ; après un an d'attente, de multiples relances et l'organisation d'une réunion entre les représentants de l'InVS et le président de la Cnil, M. Alex Türck, la Cnil répond favorablement le 17 août 2009.

1.2 Le cadre contractuel

Vingt-cinq documents, dont la signature s'est échelonnée entre février 2007 et novembre 2009, ont été nécessaires pour encadrer la réalisation de l'étude pilote, avec :

- des conventions entre l'InVS et ses partenaires, rédigées par l'InVS, qui avaient pour vocation de définir le rôle, les engagements et les modalités de fonctionnement dans le cadre de l'étude pilote,
- des contrats entre la CnamTS et chacun des partenaires de l'InVS, rédigés par la CnamTS, pour la cession des outils informatiques Foin et Camel et pour la cession à l'InVS des données ALD30 issues du Sniiram.

La production du cadre final s'est avérée extrêmement longue du fait de la double rédaction (certains documents à la charge de l'InVS, d'autres à celle de la CnamTS), du polymorphisme du cadre contractuel (conventions de partenariat, convention financière, chartes, contrats) et du nombre de parties prenantes, multipliant les échanges, les étapes de relecture et de validation des documents.
Deux contrats ont été particulièrement complexes à rédiger :

- le contrat de cession des données ALD 30 de la CnamTS à l'InVS, la CnamTS revendiquant la copropriété des données de la base croisée test, puis des travaux et publications issues des analyses de l'étude pilote (rédaction débutée le 07/01/2008, signature sept mois plus tard le 13/08/2008) ;
- la convention bi puis tripartite entre l'InVS, la CnamTS et l'Inserm, compte tenu 1/du retard de réponse de la Cnil à la deuxième demande de conseil précisant les modalités de l'exercice de retour au cas (ces modalités devant figurer dans la convention tripartite) et 2/de l'absence de personnalité morale de l'IFR69 ayant nécessité de fait de contractualiser directement avec l'Inserm. Au final, la convention tripartite, dont la rédaction a débuté le 30/04/2007 a été signée en novembre 2009 (plus de deux ans et demi après), la date de caducité étant fixée, comme l'ensemble du cadre contractuel, au 31 décembre 2009.

À la date de démarrage de la prestation informatique (juillet 2007), il est à noter qu'un certain nombre de contrats n'avaient encore pu être finalisés ni par conséquent signés. De fait, les partenaires concernés ne disposaient pas des logiciels nécessaires pour permettre leur installation sur site par le prestataire informatique.

Compte tenu de l'allongement de la durée de l'étude, l'ensemble des contrats CnamTS a du être prolongé par avenant jusqu'au 31 décembre 2009. Malgré la prestation toujours en cours à cette date, il n'a pas été possible de prolonger le cadre contractuel plus avant. En effet, une nouvelle version du logiciel Camel avait été mise en circulation dès début 2009 ; la CnamTS avait alors autorisé l'InVS à conserver la version précédente dans le cadre de la finalisation de l'étude pilote, mais le Centre de support national (CSN) ne souhaitait pas avoir à maintenir trop longtemps deux versions différentes de Camel.
Il importe de préciser que la caducité du cadre contractuel de cession de Camel impliquait automatiquement l'obligation de désinstallation et de restitution du logiciel et de sa documentation à la CnamTS.

Tableau 6 – Récapitulatif du cadre contractuel, étude pilote SMSC, 2004-2010

	Rédaction	Dénomination / partenaire(s)	Date de signature
Conventions InVS	InVS	Convention cadre de collaboration InVS / CnamTS relative à la mise en place de l'étude pilote du SMSC	15/02/2007
	InVS	Charte de fonctionnement Ile-de-France Bulletins d'adhésion pour les structures ACP pilotes (n=5)	08/11/2007 Du 10/01/2007 au 18/06/2008
	InVS	Charte de fonctionnement Nord-Pas-de-Calais Bulletins d'adhésion pour les structures ACP pilotes (n=3)	21/12/2007 Du 20/12/2007 au 18/06/2008
	InVS	Convention financière de collaboration avec le Crisapif	18/12/2007
Contrats CnamTS	CnamTS	Contrat de cession des données du Sniiram avec l'InVS	13/08/2008
	CnamTS	Contrat de cession des outils Foin et Camel avec l'InVS	20/11/2008
		Contrat de cession des outils Foin et Camel avec chacun des partenaires concernés : - prestataire informatique - ACP secteur libéral et PSPH (n=5) - Drass (n=2) - Crisapif	 12/10/2007 Du 20/12/2007 au 21/05/2008 14/04/2008 (NPC) 13/08/2008 (IDF) 13/02/2008
Convention tripartite	InVS	Convention tripartite InVS / CnamTS / Inserm (avec contrat de cession Camel en annexe)	06/11/2009

1.3 Le coût de la collaboration

1.3.1 Rémunération des ACP

La participation à l'étude des structures ACP était conditionnée à la signature d'un bulletin d'adhésion (BA) à la charte régionale SMSC Ile-de-France ou Nord-Pas-de-Calais. Le BA prévoyait une rémunération de la participation des ACP au volet rétrospectif (données 2004), compte tenu de :

- la charge de travail liée au nécessaire retour au dossier du patient ;
- la récupération *a posteriori* des variables administratives nécessaires à la génération de l'identifiant anonyme anoSNIIR1 ;
- la re-saisie spécifique des données dans un formulaire de recueil *ad hoc*.

La rémunération s'élevait à 30 € par fiche saisie et validée par l'InVS.

Il n'avait pas été prévu initialement de rémunération du volet prospectif (données 2009), l'objectif étant que les travaux avec les partenaires institutionnels puissent aboutir, courant de l'étude, à l'intégration des masques de saisie dans les logiciels métiers des ACP et à l'extraction automatique des items vers l'InVS. Cet objectif n'ayant pu être atteint pendant la réalisation de l'étude (et toujours pas à ce jour), une charge de travail identique à celle du 1er volet a été nécessaire pour la réalisation du volet prospectif. Il a donc finalement été décidé de rémunérer les ACP participants à ce second volet après attestation par l'InVS du service fait et sur présentation d'une facture. Comme pour le volet rétrospectif, la rémunération s'élevait à 30 € par fiche saisie et validée par l'InVS. Seules les structures ayant transmis les documents nécessaires ont été payées.

Pour l'ensemble des structures et les deux volets, la somme versée aux structures au 1er août 2010 s'élevait à 12 090 € (9 690 € pour le volet rétrospectif et 2 400 € pour le volet prospectif).

1.3.2 Autres coûts

Une convention financière relative à la gestion et au fonctionnement du niveau régional en Ile-de-France a été signée avec le Crisapif le 18/12/2007. Elle prévoyait une subvention allouée par l'InVS de 5 000 € pour l'achat d'un serveur dédié.

L'InVS a également procédé à l'achat du matériel nécessaire pour l'équipement des partenaires en vue de l'installation des outils Foin et Camel : achat de serveurs, de licences CFT et de lecteurs de carte à puce Camel (n=6).

2 Les travaux techniques : le développement de la solution informatique SMSC

2.1 La consultation

2.1.1 Une première consultation en novembre 2005

L'objectif initial de cette première consultation était de travailler avec les éditeurs de logiciels équipant les structures ACP pilotes en systèmes de gestion de laboratoire (SGL) afin d'intégrer le formulaire de recueil SMSC directement dans les logiciels métier et de pouvoir extraire automatiquement les CR ACP thyroïde, évitant ainsi la double saisie.
Suite aux travaux d'articulation avec la Direction générale de l'offre de soins (ex-DHOS), le Groupement pour la modernisation du système d'information hospitalier (GMSIH) et l'INCa, puis à la rencontre des éditeurs de logiciels ACP en septembre 2005, un premier cahier des charges a été préparé en 2005 et adressé, sans mise en concurrence, aux trois éditeurs de logiciels équipant les sites pilotes.
La solution fonctionnelle comprenait cinq modules pour la saisie intégrée, l'interprétation de référentiel, le stockage, l'export et la communication avec le SIH pour la récupération 1/du numéro anoSNIIR1 dans le cas des établissements publics et 2/du NIR dans le cas des structures libérales.
Pour être en cohérence avec les travaux menés au niveau international par le groupe IHE[8] sur l'échange de données ACP, l'utilisation de la norme HL7 CDA-R2 (*Health Level Seven – Clinical Document Architecture – Release 2*) était préconisée. L'application de cette norme aux données de la fiche épidémiologique thyroïde avait été préparée par le SSI en lien avec les travaux du groupe HL7 France menés dans le cadre du projet CRFS piloté par l'INCa.
Le marché s'est révélé infructueux en raison des propositions financières inacceptables au regard du faible nombre de sites concernés et de la seule localisation thyroïde, avec par ailleurs des réponses incomplètes au cahier des charges, les éditeurs semblant en attente de l'évolution attendue des systèmes d'information de santé et peu motivés par l'étude.

Tableau 7 – Première consultation marché SMSC – propositions financières des éditeurs de logiciels ACP, étude pilote SMSC, 2004-2010

Editeur	Nombre de sites à équiper	Propositions financières
X	5 sites (4 établissements, 1 structure libérale)	52 275€ HT
Y	2 sites (1 établissement, 1 structure libérale)	68 769 € HT
Z	1 site libéral	9 600 € HT
Total	**8 sites**	**130 644 € HT**

2.1.2 Une deuxième consultation en février 2007 : marché 2007-01/MULTISOURCES/AOO/DMCT

Suite à l'échec de la première consultation et à l'abandon (provisoire) des travaux de l'INCa autour des CRFS et de la norme HL7 CDA-R2, l'équipe projet a travaillé en 2006 à l'élaboration d'un deuxième cahier des charges pour le développement d'une solution complète d'anonymisation et de transfert sécurisé des données ACP vers l'InVS, cette fois-ci non intégrée aux logiciels métiers.

2.1.2.1 Formulaire de recueil des données ACP

A la demande de la direction de l'InVS, afin de limiter les coûts, le formulaire de recueil des données ACP a été développé en interne par le SSI. Le choix s'est porté sur les technologies libres (Easy PHP, serveur web Apache, base de données MySQL) pour permettre un déploiement aisé quel que soit l'équipement informatique des ACP et sans frais de licence. Non intégré au logiciel métier de l'ACP, il impliquait obligatoirement une double saisie.
Compte tenu du changement de chef de projet SSI en septembre 2006, une première version du formulaire n'a pu être adressée aux ACP pilotes qu'en juin 2007 (sur support CD-ROM), puis une seconde modifiée en septembre 2007, accompagnée d'un guide d'utilisation. L'installation a été réalisée à distance avec l'assistance de la hotline et/ou du chef de projet SSI de l'InVS. Ce formulaire a été utilisé par les ACP pour la saisie des données 2004.

2.1.2.2 Solution d'anonymisation et de remontée des données : caractéristiques principales du marché

Un appel d'offre ouvert pour « le développement d'un système d'information pour la surveillance de l'incidence des cancers thyroïdiens reposant sur le croisement de trois sources d'information » a été lancé en février 2007. A la date limite de réception des offres (avril 2007), deux sociétés avaient fait acte de candidature et ont été retenues. L'analyse des offres s'est portée en faveur de la société W ; le marché a été notifié le 11 juillet 2007 pour une durée au plus de 36 mois, dont 8 de garantie et 12 de maintenance. La réunion de lancement a eu lieu le 27 juillet 2007.

[8] Integrating the Healthcare Enterprise (IHE) est une initiative des professionnels de la santé destinée à améliorer la façon avec laquelle les logiciels du domaine échangent leurs informations. IHE propose l'utilisation coordonnée de standards établis, comme DICOM et HL7, afin d'apporter des solutions aux problèmes récurrents d'intégration de différents logiciels de santé. L'objectif est d'assurer une meilleure interopérabilité entre les systèmes qui utilisent IHE.

Le marché SMSC comprenait les prestations suivantes :

- l'intégration du module d'anonymisation (logiciels Camel et Foin fournis par la CnamTS) au niveau de l'environnement informatique des différents partenaires concernés ;
- la mise en place d'une solution sécurisée de transfert de données anonymisées au départ des structures ACP vers l'InVS *via* ses partenaires ;
- la mise en place de connecteurs informatiques entre les structures ACP des établissements publics et le SIH de l'établissement, dans le but de récupérer l'anoSNIIR1 (généré pour le PMSI) et l'appliquer aux données ACP (voir ci-dessous) ;
- la réalisation d'un outil permettant la validation des données ACP par le niveau régional du système ;
- la mise en place d'une base ORACLE pour le stockage des données à l'InVS ;
- la mise en place d'un outil de suivi des transmissions ;
- le déploiement de la solution chez les différents partenaires ;
- la formation et le transfert de compétences ;
- la maintenance corrective du système pendant un an.

Le projet était découpé en sept blocs, eux-mêmes subdivisés en 17 étapes (annexe 7).

2.2 Les choix techniques adoptés

Pour développer la solution, le prestataire s'est basé chaque fois que possible sur l'utilisation de technologies libres telles que le langage Java (plate-forme J2EE), la base de données MySql, le format de fichier XML… Certaines exigences techniques étaient connues du prestataire dès le début du projet : intégration des outils Foin et Camel dans les applications à développer aux différents niveaux du système, couplage de CAMEL avec le logiciel CFT, utilisation de la base ORACLE et de l'ETL Sunopsis à l'InVS, utilisation du format XML et de la norme NTE (fournie par la CnamTS) pour les fichiers de données à échanger.
Toutefois, certains des choix initiaux de l'InVS et/ou de la société prestataire W (fonctionnels et/ou techniques) ont dû être remis en cause du fait de contraintes (opérationnelles ou de sécurité informatique) non connues au démarrage du projet et imposées par des partenaires. Ces contraintes se sont traduites notamment par la nécessité d'abandonner la technologie web (utilisation du protocole SSH en remplacement de https, abandon des serveurs web) et d'intégrer le réseau Sintel du ministère de la Santé pour les flux régionaux et en aval du niveau régional.

2.3 Les difficultés rencontrées

2.3.1 Liées aux choix imposés par la Cnil

La procédure de récupération de l'anoSNIIR1 dans les SIH s'est révélée très complexe compte tenu de leur hétérogénéité et du manque de connaissance du circuit de l'information au sein des établissements de la plupart des personnes ressources initialement identifiées ; de fait, des difficultés à la fois organisationnelles et techniques ont été rencontrées dans la mise en place des connecteurs. Ils étaient spécifiques à chaque établissement, le plus souvent semi-manuels voire manuels, ce qui excluait l'utilisation de ce type de solution dans le cadre d'un système déployé sur les 500 sites ACP recensés sur le territoire national.
Cette problématique s'est retrouvée dans les structures privées PSPH (l'Institut Gustave Roussy et le Groupe hospitalier de l'Institut catholique de Lille), avec cette fois-ci la difficulté d'identification du lieu où était conservé le NIR afin de le rapatrier au niveau de l'application Foin installée chez le pathologiste. Des connecteurs ont également été nécessaires.

2.3.2 Liées aux choix imposés par la CnamTS

- **Les outils Foin et Camel**

L'utilisation de Foin par l'InVS et ses partenaires était conditionnée à une livraison par la CnamTS d'une version « encapsulée » dans le logiciel Camel (sauf dans le cas des structures ACP, pour lesquelles les contraintes techniques liées à l'utilisation de Camel étaient inenvisageables).
L'utilisation de Camel chez les partenaires régionaux et nationaux était extrêmement contraignante. En plus des contraintes juridiques (cession liée obligatoirement à un contrat par site hébergeur), son installation nécessitait la mise à disposition d'un serveur dédié, non web. Le couplage avec un logiciel assurant l'automatisation des transferts tel CFT était une contrainte technique supplémentaire. En effet, CFT est un outil relativement ancien qui n'était plus commercialisé, au moment de l'étude, que par une seule société. L'InVS a dû procéder à l'achat de licences avec maintenance pour l'ensemble des partenaires. Or, au moment de l'achat, seule une version plus récente que celle préconisée dans la documentation Camel, fournie par la CnamTS, était encore commercialisée. Après vérification avec la CnamTS, cette version était annoncée comme compatible avec Camel.

En pratique, la société prestataire W a rencontré des difficultés dans la mise en place des flux Camel-CFT avec un flux qui tombait parfois subitement, une désynchronisation spontanée entre Camel et CFT et dans certains cas, une impossibilité de mettre en place le flux (cas de la liaison Drass Nord-Pas-de-Calais / InVS qui n'a jamais fonctionné). La question de la compatibilité reste posée d'autant plus que la CnamTS, informée de ces difficultés, a rapporté des situations semblables pour certaines caisses primaires d'assurance maladie, utilisant la même version de CFT que l'InVS pour leurs transmissions de données.
À noter que la mise à disposition de Foin dans le cadre de l'étude pilote SMSC a nécessité pour la CnamTS la passation d'un marché pour le développement du module en Java afin de faciliter son installation au niveau d'environnements informatiques hétérogènes.

- **La problématique réseau**

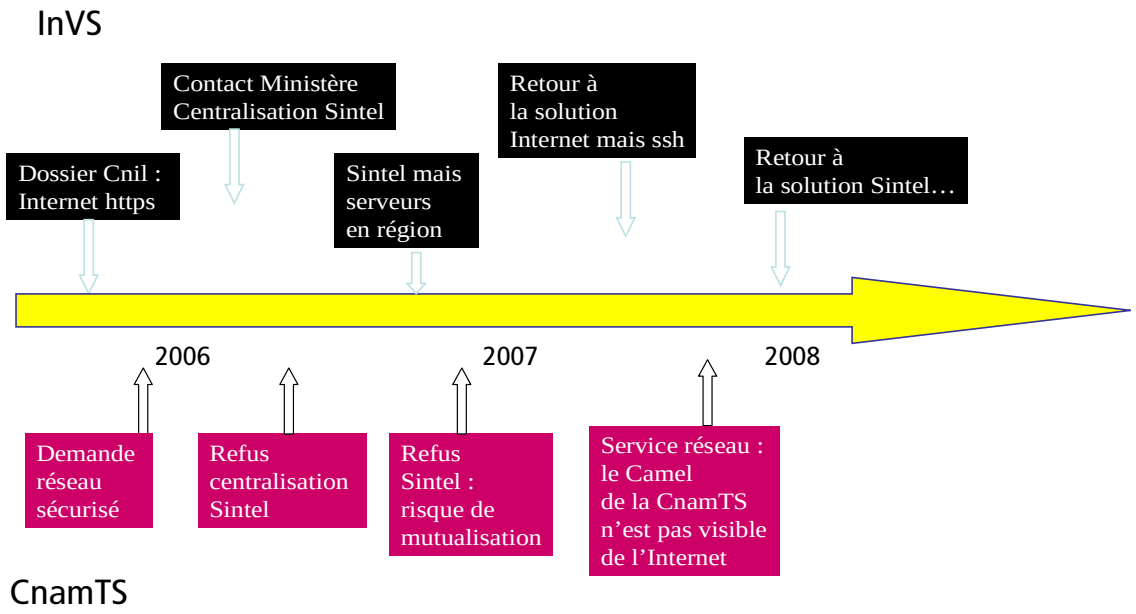
La question de la circulation des flux dans le cadre de l'étude a été extrêmement complexe, du fait du transit des données ACP :
1/ par les Drass, situées sur le réseau Sintel du ministère chargé de la Santé ;
2/ par le Centi de la CnamTS, situé sur le réseau Ramage, réseaux tous deux très sécurisés.

Le cahier des charges de l'InVS prévoyait que les serveurs Drass soient physiquement installés dans les Drass et que les flux de données entre les partenaires, notamment à l'arrivée et au départ du niveau régional, circulent par internet sécurisé (protocole https prévu dans le dossier Cnil).
Fin 2006, dans le cadre de la finalisation du cahier des charges informatique, l'InVS a rencontré les responsables des services informatiques (CROMI) des deux Drass concernées, puis les représentants de Sintel au ministère de la Santé. La proposition de centraliser les flux régionaux au niveau du ministère (dans des bases régionales distinctes), dans la mesure où il n'y avait pas d'intervention humaine au niveau régional sur les données identifiantes, a été émise par Sintel. Outre qu'elle ne modifiait en rien les sécurités informatiques, cette solution présentait l'intérêt pour l'InVS d'avoir un interlocuteur unique lors de la phase de généralisation du SMSC.
Sollicitée fin 2006 sur cette proposition d'organisation, la CnamTS avait émis un certain nombre de réserves et demandé, dans l'hypothèse d'une mise en place effective, une autorisation Cnil. Devant ces réticences, l'InVS était alors revenu à son projet initial de transit des données par internet sécurisé, les serveurs restant physiquement dans les Drass.

Cette problématique a ressurgi en octobre 2007 au moment de l'ouverture effective des flux par la société prestataire W. La CnamTS a informé l'InVS que le serveur Camel du Centi, recevant les flux entrants des deux Drass, n'était pas visible d'internet[9]. Elle a donc demandé à l'InVS d'étudier un nouveau schéma des flux utilisant 1/ soit la liaison réseau existante entre le ministère de la Santé et la CnamTS ou 2/ un réseau protégé spécifique à créer *de novo*.
Compte tenu des contraintes de calendrier liées à la prestation en cours, la deuxième solution a été rapidement écartée et des contacts ont été pris dès janvier 2008 avec le service Sintel 4 du ministère de la Santé pour examiner à nouveau une solution utilisant le réseau Sintel.
Plusieurs réunions entre l'InVS, Sintel, la CnamTS, les Drass, le tiers de confiance et la société prestataire W ont été nécessaires pour valider le schéma final de circulation des flux (annexe 8) ; celui-ci prévoyait la réception des flux par Sintel, la CnamTS et l'InVS sur des serveurs SAS situés hors des réseaux protégés respectifs (en zone démilitarisée ou DMZ), à mettre en place *de novo*.
Cette organisation s'inscrivait d'emblée dans une démarche pérenne de probable généralisation à l'issue de l'étude pilote. La solution alternative de transmission des données entre les partenaires par CD-Rom, proposée par la CnamTS, a été refusée par Sintel et était inenvisageable dans l'hypothèse d'un système déployé.

[9] L'InVS a également été contraint d'abandonner le choix des technologies libres (largement basées sur le web) que ce soit pour le formulaire de recueil que pour le reste de la solution (voir 2.4.3 Calendrier de la prestation, avenants 1 et 2)

Figure 4 : Problématique d'interconnexion réseau, étude pilote SMSC, 2004-2010
La figure ci-dessous illustre la multiplicité et les difficultés d'échanges avec la CnamTS sur la problématique réseau.


InVS
Contact Ministère Centralisation Sintel
Retour à la solution Internet mais ssh
Dossier Cnil : Internet https
Sintel mais serveurs en région
Retour à la solution Sintel…
2006
2007
2008
Demande réseau sécurisé
Refus centralisation Sintel
Refus Sintel : risque de mutualisation
Service réseau : le Camel de la CnamTS n'est pas visible de l'Internet
CnamTS

1

Après accord de principe de la CnamTS, la faisabilité technique de la solution a été confirmée en mai 2008 pour une livraison prévisionnelle fin septembre 2008. Courant de l'été 2008, l'InVS a fait appel à un transporteur pour assurer le déménagement des deux serveurs Drass vers la salle serveur du ministère de la Santé.
Finalement, la livraison d'une documentation par la CnamTS n'a pu être effectuée qu'en février 2009.
L'InVS a présenté pour information l'architecture finalement retenue à la Cnil en août 2008 dans le cadre de la 2e demande de conseil.

2.4 Le suivi de la prestation

2.4.1 Instances de pilotage et de suivi

Pour la société prestataire W, le projet était suivi par un chef de projet qui a changé en cours de prestation (départ en décembre 2008).
Trois instances de pilotage et de suivi ont été mises en place, l'InVS assurant la MOA du projet et la société prestataire W, la MOE.

- Un comité de pilotage, chargé de vérifier l'avancement des travaux et de gérer d'éventuels incidents. Il était prévu de le réunir à chaque fin de bloc (voir le phasage du projet en annexe 7), afin de valider les travaux effectués. En pratique, le Copil n'a pu être réuni qu'une seule fois, à la fin du bloc 1, sanctionné par la livraison des spécifications fonctionnelles et techniques. Dans la suite de la prestation, compte tenu des difficultés rencontrées et des retards accumulés, il n'a pas été possible pour le prestataire de réaliser les blocs les uns après les autres : le développement des différentes étapes et blocs a du être mené en parallèle.
- Un comité de suivi, composé principalement de l'équipe projet DMCT – SSI et du chef de projet de la société prestataire W, chargé de vérifier l'avancement détaillé des travaux, de résoudre les difficultés techniques et de gérer les éventuels incidents. Il s'est réuni en moyenne tous les 15 jours jusqu'au départ de la responsable du programme SMSC en avril 2009.
- Un comité de gestion interne, qui réunissait en comité restreint les deux chargés de projet SMSC au DMCT et au SSI et le chef de projet de la société prestataire W. Durant toute la durée de la prestation, il s'est réuni de façon hebdomadaire le plus souvent dans le cadre de réunions téléphoniques. Il était chargé du suivi et de l'analyse au plus prêt de l'avancement du projet, de définir les tâches à venir et de résoudre les problèmes de toute nature pouvant survenir avec les partenaires.

Les comptes-rendus de l'ensemble de ces réunions devaient être rédigés par la société W. Initialement livrés sous forme de RIDA (relevés d'information, décisions, actions), ils se sont espacés au fil du temps pour disparaître quasi-complètement en fin de prestation malgré les demandes réitérées de l'équipe projet, qui s'est trouvée dans la nécessité d'en assurer la rédaction. Il est à noter que ce manque d'outil de suivi opérationnel a été préjudiciable au suivi du projet en particulier pendant les derniers mois, avec une perte d'information et de visibilité sur l'état d'avancement des travaux engagés.
Outre ces instances de pilotage et de suivi de la prestation, deux réunions de point d'avancement ont été organisées avec les partenaires participant au circuit de remontée des données ACP vers l'InVS (pathologistes et informaticiens des structures ACP pilotes, Drass, Crisapif, tiers de confiance). Ces réunions ont ensuite été interrompues compte tenu du ralentissement du rythme des développements et des problématiques rencontrées en 2008, notamment autour de l'introduction dans le circuit du réseau Sintel du ministère de la Santé.

De façon moins formalisée, de nombreuses réunions ont été organisées pendant tout le déroulement de la prestation entre l'InVS, la société prestataire W et les partenaires du projet, en particulier la CnamTS et les services Sintel du ministère de la santé.

2.4.2 Gestion des non-conformités

Une application Mantis[10] dédiée au suivi des anomalies a été mise en place en février 2008 par la société prestataire W sur ses serveurs, à la demande de l'InVS.
Les dysfonctionnements ont été enregistrés par l'équipe projet InVS tout d'abord sur l'application de recueil des fiches ACP, puis sur les autres applications du projet :

- CRACP : recueil fiches ACP ;
- ACPService : remontées des fiches ACP vers le niveau régional ;
- CRACPRef : contrôle qualité des fiches au niveau régional ;
- ACPRefService : réception des fiches médicales ACP et envoi vers le niveau national ;
- DRASSService : réception des fiches identifiantes ACP et envoi vers la CnamTS.

2.4.3 Calendrier de la prestation

Le calendrier réalisé est présenté ci-dessous. Au total, le projet a été prolongé de 15 mois par rapport au planning initial, soit une prestation étalée sur 51 mois au lieu des 36 initialement prévus.
Trois avenants ont été nécessaires.

- **Avenant n°1 (24/12/2007)**

L'objet du premier avenant était :

- de décaler la livraison du volet « connecteurs entre structure ACP et SIH » devant figurer dans les spécifications fonctionnelles et techniques, sans prolonger la durée globale du marché ni modifier substantiellement le planning ;
- de confier au prestataire le développement d'un nouvel outil pour la saisie des fiches thyroïde, en lieu et place du formulaire InVS. Ce nouveau formulaire de recueil devait comprendre des contrôles de cohérence (contrôle du NIR par sa clé) et la génération d'un compte-rendu ACP pouvant être imprimé et utilisé par le pathologiste déclarant.

Le décalage de la livraison du volet relatif aux connecteurs faisait suite à la rencontre de plusieurs difficultés liées :

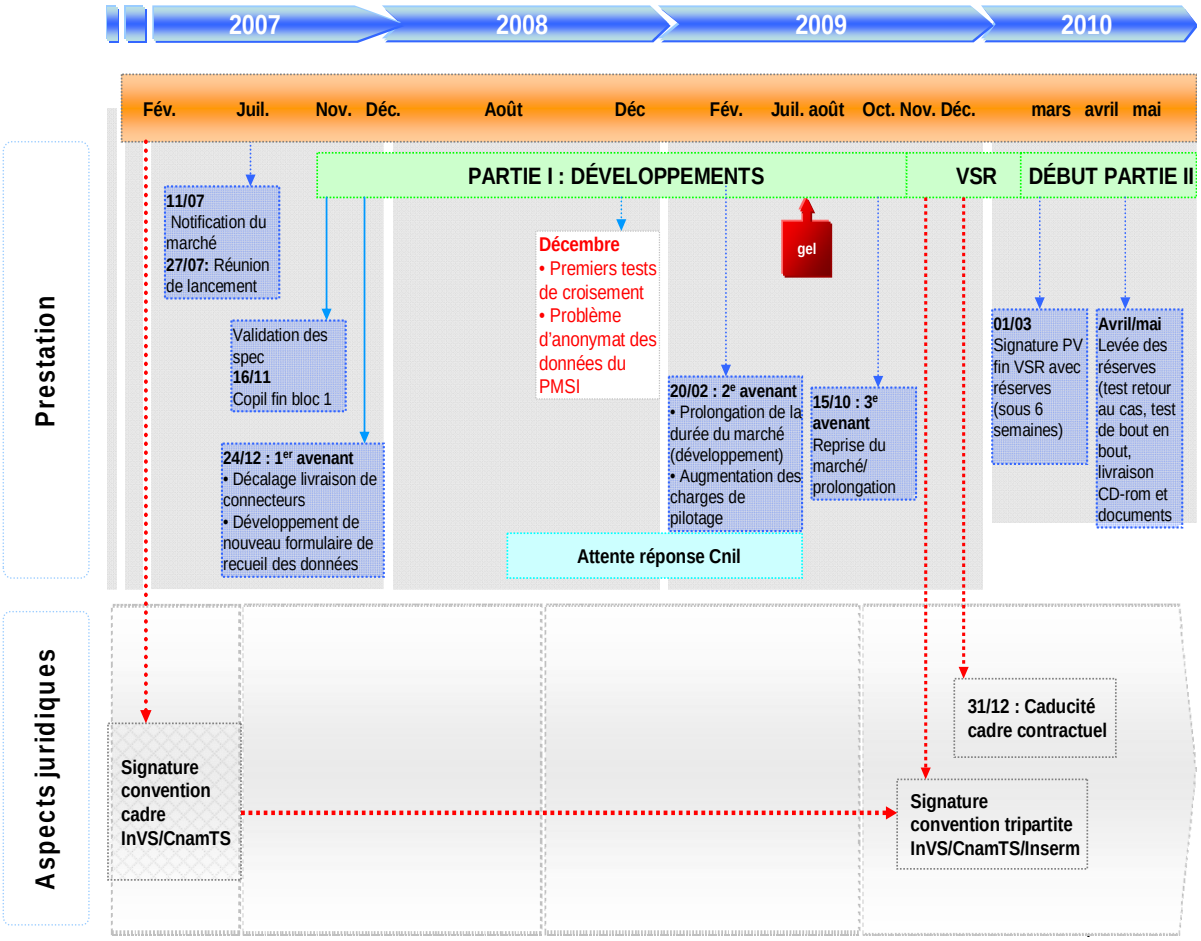
- à l'identification de la personne ressource en charge des SIH au sein des établissements concernés ;
- au manque de disponibilité de la personne ressource une fois identifiée ainsi que des techniciens ;
- au manque de connaissance du circuit et du lieu de stockage des éléments administratifs d'identification du patient (numéro anoSNIIR1 du PMSI ou NIR dans le cas des établissements privés PSPH – IGR et GHICL) par les personnes ressources.

La commande d'un nouveau formulaire de recueil faisait suite à la découverte, à la lecture du premier contrat de cession Foin rédigé par la CnamTS et adressé à la société prestataire W en septembre 2007, que le module Foin (de même que le logiciel Camel) ne pouvait être installé sur un serveur web. Or le formulaire développé par le SSI et transmis aux structures ACP pour le recueil de données utilisait la technologie web (serveur web Apache) et ne pouvait donc être utilisé. Par ailleurs, il ne permettait pas la génération automatique d'un compte-rendu, fonctionnalité souhaitée par

[10] Mantis est un système de suivi d'anomalies logicielles (bugs) basé sur une interface web. Le principe de cet outil consiste à enregistrer la déclaration d'un bug informatique, puis, pour les techniciens de maintenance informatique concernés, à mettre à jour l'avancement de sa résolution, jusqu'à sa clôture. Le déclarant de l'anomalie peut s'informer à tout moment via le serveur Web de l'avancement du traitement de son problème. Pour plus d'information, consulter le site http://www.mantisbt.org/

certains pathologistes. Ce nouvel outil a donc été pensé de telle sorte à pouvoir permettre à l'ACP de générer un compte-rendu en cas d'utilisation du formulaire dans sa pratique de routine, l'idée étant de tester son ergonomie et son utilité dans le cadre de la réalisation du volet prospectif (objectif : proposer à terme cet outil aux ACP non équipés de logiciel métier).
L'avenant 1 représentait une augmentation d'un peu plus de 9 % du montant du marché initial.

Figure 5 : Frise chronologique de la prestation SMSC



- **Avenant n°2 (20/02/2009)**

Le deuxième avenant avait pour objet de prolonger la durée de la partie « développements » (donc la durée globale du marché) et d'intégrer une augmentation des charges de pilotage du projet. Le délai d'exécution pour la partie « développements » a été fixé à 22 mois au plus au lieu des 16 prévus initialement.
La passation de l'avenant n°2, demandée par la société prestataire W, était liée à différents aléas et retards peu voire non anticipables et survenus pour la majorité d'entre eux chez des partenaires tiers du projet, avec pour conséquences :

- un retard au démarrage de certaines étapes (le prestataire n'a pas toujours pu disposer des pré-requis nécessaires en temps utile) ;
- la mise en place d'infrastructures non prévues initialement et motivées par des contraintes (opérationnelles ou de sécurité informatique) non connues au démarrage du projet ;
- des remises en question de paramétrages ou de développements déjà effectués, avec nécessité de retours (parfois multiples) sur les applications installées chez les partenaires ;
- un allongement de la durée du projet avec dérive des charges de pilotage pour la société.

Les avenants 1 et 2 représentaient une augmentation de 15 % du montant initial du marché (maximum autorisé par le code des marchés publics), portant le montant global de la prestation de 128 535,00 € HT à 135 164,60 € HT.

- **Avenant n°3 (10/2009)**

L'objet de ce dernier avenant était une nouvelle prolongation de la durée de la partie « développements » suite à une période de suspension des prestations (gel en juillet et août 2009 consécutif au retard de réponse de la Cnil sur les modalités d'exercice de retour au cas, entraînant un arrêt des travaux de la société prestataire W chez le tiers de confiance).
Dans ce cadre, la durée globale du marché a été étendue à 51 mois, avec une fin de la période de vérification de service régulier fixée au 28/02/2010. Un procès verbal de fin de VSR (avec réserves à lever sous six semaines) a effectivement été signé le 1er mars 2010, officialisant la fin de l'étude pilote SMSC.

3 Les travaux institutionnels connexes

En parallèle à l'avancée des travaux autour de la réalisation de l'étude pilote SMSC, l'InVS a suivi l'évolution des projets connexes basés sur l'ACP et pilotés par l'INCa en lien avec les instances de la profession.
Dans ce cadre, l'équipe projet a été amenée à participer à différents groupes de travail :

- Groupe INCa / Direction générale de l'offre de soins / InVS en 2005 (jusqu'à début 2006, date à laquelle la DGOS s'est retirée du groupe et les travaux autour de la norme HL7 CDA-R2 ont été abandonnés temporairement par l'INCa) ;
- Groupe INCa / InVS dans le cadre du suivi de la prestation passée par l'INCa avec la société Santéos pour la mise en place des CRFS (groupe piloté par le Pr Voigt en 2007 et 2008), qui a abouti à la production de CRFS par localisation cancéreuse. Les CRFS constituent un référentiel mais ont été jugés « trop expertaux » par la profession ACP ;
- Groupe INCa/instances représentatives ACP (Société française de pathologie - SFP, Association française d'assurance qualité en ACP - AFAQAP et syndicat des médecins pathologistes français - SMPF) pour la définition d'items minimaux opposables par localisation cancéreuse, version simplifiée et consensuelle des CRFS dans le cadre de la procédure d'agrément des établissements de soins prenant en charge des patients atteints de cancer (depuis 2008).

IV – RESULTATS DE L'ANALYSE DES DONNÉES CROISÉES

1 Analyse des cas rétrospectifs de 2004

1.1 La base ACP

Les huit structures pilotes ont participé au volet rétrospectif portant sur le recueil des données 2004. Le décompte agrégé de la transmission des fiches (tableau 8) montre qu'au total, 453 fiches ont été saisies parmi lesquelles 442 ont été validées (97 %).
Les cas de non-validation (n=11) concernaient principalement les structures libérales (n=8 soit 73 % des cas) et étaient dus à un échec de récupération du NIR du patient (6 cas) et à une erreur de contrôle du NIR par sa clé (2 cas). Les échecs de récupération du NIR étaient dus à la fermeture de la clinique ayant adressé le prélèvement ou à l'archivage des dossiers correspondants par la clinique émettrice qui ne pouvait pas les mettre à la disposition du projet. Parmi les établissements participants, le CHRU de Lille n'a pu valider 3 fiches saisies.
En raison d'erreur de transmission portant sur 11 autres fiches, seulement 431 des 442 fiches validées ont été transmises au niveau régional. Ces erreurs de transmission concernaient exclusivement les établissements de soins et étaient liées dans tous les cas à des problèmes de récupération du numéro anonyme Foin1 dans le SIH.
L'analyse a finalement porté sur 423 fiches de données individuelles obtenues après exclusion de 3 doublons et de 5 fiches dont la date d'enregistrement n'était pas l'année 2004. Sur les 423 fiches, 250 provenaient de la région Ile-de-France et 173 du Nord-Pas-de-Calais.
Le pourcentage des informations renseignées et exploitables (tableau 9) variait de 53 % à 100 %. Les variables les moins renseignées étaient le lieu de résidence pour lequel on ne disposait que de 23 % d'informations en Ile-de-France.

Tableau 8 – Bilan par site de la transmission agrégée des données ACP 2004

	Nombre de fiches saisies	Erreur de validation ou de transmission	Nombre de fiches validées	Erreur d'émission	Nombre de fiches transmises au niveau régional
RÉGION IDF					
Laboratoire de Tolbiac	44	3	41	0	41
Institut de pathologie de Paris	5	2	3	0	3
APHP Pitié Salpétrière	122	0	122	3	119
Institut Gustave Roussy	79	0	79	2	77
APHP Ambroise Paré	20	O	20	3	17
Sous-total IDF	270	5	265	8	257
RÉGION NPdC					
Laboratoire de Valenciennes	20	3	17	0	17
Groupe hospitalier de l'Institut catholique de Lille	2	0	2	0	2
CHRU Lille	161	3	158	3	155
Sous-total NPdC	183	6	177	3	174
TOTAL	453	11	442	11	431

Tableau 9 : Pourcentage d'informations recueillies et exploitables dans la base ACP, 2004 et description des cas

Variables	% informations renseignées et exploitables	Résultats N (%)
Données administratives (N = 423)		
Sexe	100	
- Masculin		99 (23)
- Féminin		324 (77)
Date de Naissance	100	
Âge (en année)	100	
- < 20		8 (2)
- 20-29		38 (9)
- 30-39		70 (17)
- 40-49		85 (20)
- 50-59		104 (25)
- 60-69		74 (17)
- 70 et +		44 (10)
Code postal	54	
- Ile-de-France		58 (23)
- Nord-Pas-de-Calais		172 (99)
Commune de domicile	53	
Date d'enregistrement	100	
Données sur la pièce opératoire (N = 423)		
Nature de la pièce opératoire	99	
- Lobectomie, loboisthmectomie, isthmectomie		54 (13)
- Thyroidectomie totale		282 (67)
- Thyroïdectomie sub totale		60 (14)
- Biopsie thyroïdienne, autre		25 (6)
Poids de la pièce opératoire en gramme	92	
- ≤ 20		131 (34)
- entre 20 et 50		180 (46)
- > 50		78 (20)
Taille lobe droit	85	
Taille lobe gauche	82	
Nombre total de prélèvement sur la thyroïde examiné	77	
Nombre total de prélèvement sur la lésion	69	
Geste ganglionnaire	97	
- Curage		116 (28)
- Picking		58 (14)
- Aucun		237 (58)
Données sur la lésion (N = 423)		
Taille de la lésion la plus grande en mm	98	
- ≤ 4		122 (30)
- entre 4,1 et 10		107 (26)
- entre 10,1 et 20		90 (22)
- > 20		91 (22)
Nombre de foyer de lésions	98	
- Unifocal		285 (69)
- Multi focal		131 (31)
Capsule de la lésion	82	
- Encapsulée		132 (33)
- Non encapsulée		237 (67)
Lésions encapsulées (variable conditionnelle, n = 108)	89	
- Non infiltrées		42 (39)
- Infiltrées non dépassées		25 (23)
- Dépassées		41 (38)

Invasions vasculaires	65	
- Non		216 (79)
- Douteux		8 (3)
- Oui manifeste		49 (18)
Invasions vasculaires manifeste $\geq$ 4 (variable conditionnelle, n = 49)	90	
- Oui		14 (32)
- Non		30 (68)
Invasions extra thyroïdiennes	94	
- Non		337 (84)
- Oui T3		56 (14)
- Oui T4		7 (2)
Nombre de ganglions	47	
- 0		6 (3)
- entre 1 et 5		55 (27)
- entre 6 et 10		37 (18)
- supérieur à 10 (maximum = 97)		103 (51)
pT	93	
- Taille de la tumeur limitée à la glande thyroïde $\leq$ 2cm		276 (70)
- Taille de la tumeur limitée à la glande thyroïde (2-4 cm)		46 (12)
- Taille de la tumeur limitée à la glande thyroïde, > 4 cm		69 (17)
- Tumeur dépassant la capsule		4 (1)
pN	93	
- Pas de ganglions régionaux métastatiques		138 (35)
- Ganglions lymphatiques régionaux homolatéraux		6 (2)
- Ganglions para trachéaux, trachéaux ou pré laryngés		15 (4)
- Ganglions autres chaines cervicales, médiastinaux		39 (10)
- Renseignements insuffisants		195 (49)
Métastase	85	
- Présence		1 (0)
- Absence		42 (12)
- Non disponible		316 (88)
Type histologique	100	
- Vésiculaire		40 (9)
- Papillaire ou folliculaire		349 (82)
- Médullaire		18 (4)
- Autre		16 (4)
Lésions associées		
- Lésions associées 1	93	
- Lésions associées 2	66	
- Lésions associées 3	67	

L'âge moyen des cas au diagnostic était de 50,7±15,7 ans chez les hommes et de 49,4±15,2 ans chez les femmes qui représentaient plus de 75 % des cas. La forme histologique la plus fréquente était la forme papillaire (82 %) qui était plus fréquente dans le Nord-Pas-de-Calais (89 %) qu'en Ile-de-France (78 %). D'une manière générale les lésions de petite taille inférieure à 4 mm étaient prédominantes (30 %).

1.2 Les bases PMSI et ALD30 2004

Un total de 2 947 cas de cancer de la thyroïde, correspondant à 5 018 séjours ont été identifiés dans la base PMSI après suppression de 4 % d'identifiants non valides (non exploitables). Parmi ces 2 947 cas repérés grâce à l'algorithme de sélection, 2 240 (76 %) étaient domiciliés en Ile-de-France et 694 (24 %) dans le Nord-Pas-de-Calais.
Quant à la base ALD30 du Sniiram, elle contenait 1 159 cas de cancer de la thyroïde dont 921 (79 %) étaient domiciliés en Ile-de-France et 238 (21 %) dans le Nord-Pas-de-Calais.
Parmi les 1 159 patients en ALD, 202 étaient retrouvés parmi les données du PMSI ou ACP et provenaient des régimes d'Assurance maladie suivants : régime général, régime social des indépendants, régime des militaires de carrière (tableau 10).

Tableau 10 : Origine des cas de cancer de la thyroïde mis en ALD appariés avec les données du PMSI ou des ACP

Régime d'Assurance maladie	Nombre (%)
Régime général	189 (94)
Régime social des indépendants	10 (5)
Régime des militaires de carrière	1(0)
Autres	2 (1)
Total	202 (100)

1.3 Nombre de patients communs identifiés dans les différentes bases croisées

En croisant les bases PMSI (2 947 patients) et ALD30 (1 159 patients), nous avons obtenu 749 patients appariés que nous annoterons $(PMSI/ALD)^+$ dans la suite du document.
Le croisement des données ACP avec les 749 données $(PMSI/ALD)^+$ a identifié 191 (45 %) cas communs aux trois bases que nous désignerons $(PMSI/ALD/ACP)^+$**.**
Les effectifs communs aux données ACP croisées avec les données du PMSI étaient de 343 (81 %) et de 202 (48 %) pour les données ACP croisées avec les données ALD30. Dans la suite du document ces croisements seront respectivement annotés $(PMSI/ACP)^+ ALD^-$ et $(ALD/ACP)^+ PMSI^-$ (tableaux 11, 12 et figure 5).

Tableau 11 : Résultat du croisement des données de la base ACP avec les données PMSI et ALD30

	Base ACP (N = 423)	
	Retrouvés (%)	Non retrouvés (%)
PMSI (N= 2 947)	343 (81)	80 (19)
ALD30 (N = 1 159)	202 (48)	221 (52)
$(PMSI/ALD)^+$ (N= 749)	191 (45)	232 (55)

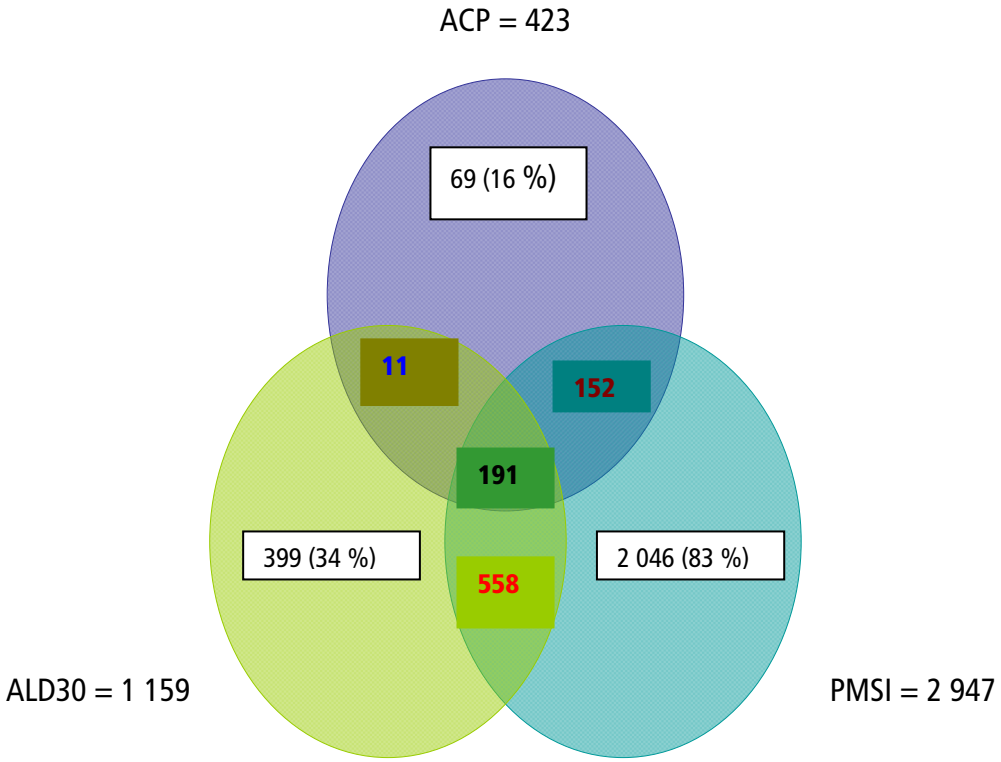
Parmi les 232 données ACP non présentes dans la base $(PMSI/ALD)^+$, 152 étaient retrouvés dans la base PMSI et 11 dans la base ALD30 (figure 5).
La base finale d'analyse est décrite dans le tableau 12.

Tableau 12 : Composition de la base d'analyse

Bases croisées	Effectif apparié (%)
$(PMSI/ALD/ACP)^+$	191 (45)
$(PMSI/ACP)^+ALD^-$	152 (36)
$(ALD/ACP)^+ PMSI^-$	11 (3)
$ACP^+ALD^-PMSI^-$	69 (16)
Base d'analyse	**423**

Le nombre de données ACP retrouvées dans la base PMSI était de 343 (191 + 152), tandis que celui retrouvé dans les ALD était de 202 (191 + 11). Le nombre de données ACP retrouvées dans la base PMSI ou dans la base ALD30 était de 354 (191 + 152 + 11).

Figure 6 : Répartition des effectifs appariés selon les sources



1.4 Niveau de cohérence des données appariées

1.4.1 Niveau de cohérence des données appariées entre PMSI et ALD (tableau 13)

La comparaison individuelle des 749 données du (PMS/ALD)$^{+}$ entre les bases PMSI et ALD montrait une cohérence parfaite sur le sexe et le département de domicile. Un seul cas identifié comme sexe masculin parmi les données ALD est renseigné comme sexe féminin dans la base PMSI.
En ce qui concerne le département de domiciliation, dans la région du Nord-Pas-de-Calais, une cohérence de 100 % a été observée entre les données ALD et PMSI. En Ile-de-France elle variait de 97 % en Seine-Saint-Denis à 99 % à Paris, dans l'Essonne, en Seine et Marne et dans les Yvelines.
La moyenne d'âge était comparable entre les deux sources. Les données ALD surestimaient de 1 an l'âge de 38 % des patients de la base PMSI. Ceci est dû au fait que l'âge dans la base PMSI était renseigné en âge révolu et celui des ALD en âge atteint. Lorsque cet écart était pris en compte dans l'analyse, on observait une cohérence totale sur l'âge entre les deux sources.

Tableau 13 : Niveau de cohérence des données entre le PMSI et les ALD (n=749)

Variables	Effectif ALD (n=749)	Effectif PMSI (n=749)	% de cohérence
Sexe			99
Masculin	168	167	
Féminin	581	582	
Âge moyen	48,5	48,1	100[a]
Région NPdC[b]			
Nord	85	85	100
Pas-de-Calais	66	66	100
Région IDF[b]			
Paris	127	127	99
Seine-et-Marne	58	59	99
Yvelines	89	89	99
Essonne	74	75	99
Hauts-de-Seine	73	74	98
Seine-Saint-Denis	60	58	97
Marne	55	55	98
Val d'Oise	57	56	98

[a] Cohérence calculée sur l'âge des cas ; [b]Les données manquantes (5) ont été exclues dans le calcul de pourcentage de cohérence sur la résidence

1.4.2 Niveau de cohérence des données appariés : (PMSI/ALD/ACP)+ (n=191)

On notait une cohérence de 99 % du sexe et de 100 % de l'âge des cas.
Compte tenu du fait que dans la base ACP, seulement 23 % des données concernant la résidence des patients étaient exploitables dans la région Ile-de-France et 99 % dans le Nord-Pas-de-Calais, la recherche de cohérence sur la résidence n'a été effectuée que dans le Nord-Pas-de-Calais pour cette base. Ainsi parmi les 191 patients communs aux trois bases, le département de domiciliation des 67 patients du Nord-Pas-de-Calais était cohérent (tableau 14).

Tableau 14 : Niveau de cohérence entre les données croisées (PMSI/ALD/ACP)+

	(PMSI/ALD/ACP)+ (n=191)			
Variables	**Effectif ACP**	**Effectif PMSI**	**Effectif ALD**	**% de cohérence**
Sexe				99
Masculin	40	39	40	
Féminin	151	152	151	
Âge moyen	49	48,6	49	100 [a]
Région NPdC				
Nord	46 [b]	47	47	100
Pas de Calais	21	21	21	100
Région IDF				
Paris	-	38	39	-
Seine-et-Marne	-	11	11	-
Yvelines	-	17	17	-
Essonne	-	14	14	-
Hauts-de-Seine	-	18	17	-
Seine-Saint-Denis	-	5	6	-
Marne	-	15	14	-
Val d'Oise	-	5	5	-

[a] Cohérence calculée sur l'âge des cas ; [b] Donnée manquante

1.4.3 Niveau de cohérence des données appariés : (PMSI/ACP)+ (n=343)

La recherche du niveau de cohérence individuelle entre les données PMSI et ACP concerne 343 patients communs aux deux bases (tableau 15).
Une cohérence de 100 % a été respectivement observée sur le sexe et sur l'âge.
Pour les mêmes raisons évoquées dans le paragraphe ci-dessus (1.4.2), la recherche du niveau de cohérence entre les observations sur la domiciliation n'a été réalisée que sur les départements du Nord-Pas-de-Calais.

Ainsi on a observé 100 % de cohérence sur le département de domiciliation (Nord) entre les données ACP et PMSI hormis le cas d'un patient répertorié dans la base PMSI comme résidant dans le Nord mais renseigné comme résidant dans le Pas de Calais par les données ACP. En ce qui concerne l'âge, seulement 59 % des données étaient cohérentes. Cette cohérence de l'âge atteint 100 % lorsqu'on prenait en compte l'écart entre l'âge renseigné en âge révolu (PMSI) et atteint (ACP).

1.4.4 Niveau de cohérence des données appariées : (ALD/ACP)+ (n=202)

La recherche du niveau de cohérence individuelle entre les données ALD et ACP concerne 202 patients communs aux deux bases. Un pourcentage de 100 % des données sur le sexe, l'âge et la résidence étaient cohérentes (tableau 15).

Tableau 15 : Niveau de cohérence entre les données ACP, PMSI et ALD croisées deux à deux

	(PMSI/ACP)+ (n=343)			(ALD/ACP)+ (n=202)		
Variables	**ACP**	**PMSI**	**% de cohérence**	**ACP**	**ALD**	**% de cohérence**
Sexe			99			100
Masculin	79	79		41	41	
Féminin	264	264		161	161	
Âge moyen	49,8	49,3	100	48,9	48,9	100 [a]
Région NPdC						
Nord	105 [b]	107	100	47	48	100
Pas-de-Calais	47	46	98	21	21	100

[a] Cohérence calculée sur l'âge des cas ; [b] Donnée manquante

1.5 Taux d'appariement

1.5.1 Définition

Le taux d'appariement des données croisées des trois sources est égal au nombre de cas communs aux bases croisées/Nombre total de cas ACP.
Ces taux peuvent se décliner par département et/ou classe d'âge

1.5.2 Taux d'appariement des données selon le type de croisement (tableau 16)

Le taux d'appariement des trois sources s'élevait à 84 %. Il variait selon la région, 89 % dans le Nord-Pas-de-Calais et 80 % en Ile-de-France et selon l'âge de 63 % avant 20 ans à 90 % chez les 30-39 ans.
Le plus faible taux d'appariement (48 %) était observé entre les données ALD et ACP, tandis que le croisement des sources PMSI/ACP montrait un taux d'appariement de 81 % proche de celui des trois sources croisées.

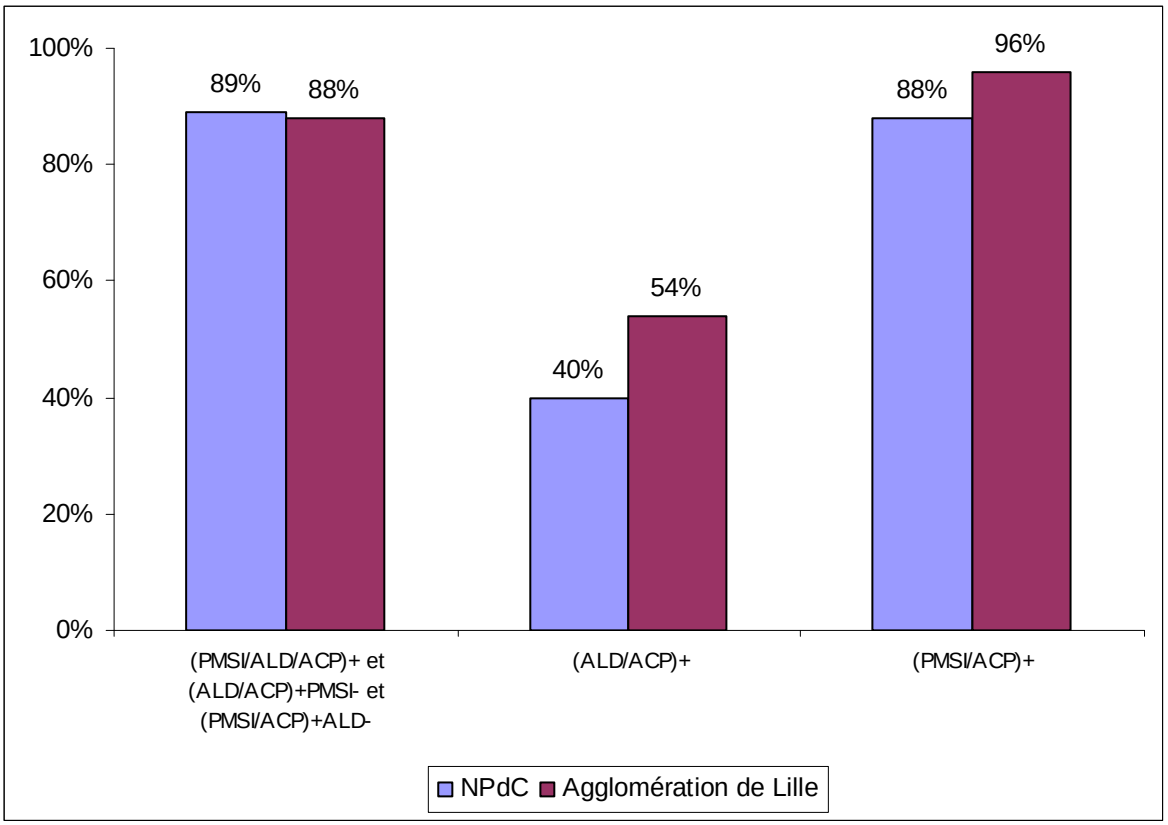
Tableau 16 : Effectif et taux d'appariement selon le type de croisement

	ACP	**(PMSI/ALD/ACP)+ et (PMSI/ACP)+ALD- et (ALD/ACP)+PMSI-**		**(ALD/ACP)+**		**(PMSI/ACP)+**	
	Effectif	Effectif	Taux en %	Effectif	Taux en %	Effectif	Taux en %
Total	**423**	**354**	**84**	**202**	**48**	**343**	**81**
Selon la région							
NPdC	173	154	89	69	40	153	88
IDF	250	200	80	133	53	190	76
Selon l'âge (an)							
< 20	8	5	63	0	-	5	63
20-29	38	34	89	22	58	32	84
30-39	70	63	90	42	60	62	89
40-49	85	65	76	40	47	63	74
50-59	104	87	84	44	42	85	82
60-69	74	61	82	32	43	57	77
70 et plus	44	39	89	22	50	39	89

1.5.3 Comparaison des taux d'appariement du Nord-Pas-de-Calais avec les taux du registre de l'agglomération de Lille de 2005

En comparant les taux d'appariement des données de la région du Nord-Pas-de-Calais de notre étude avec ceux obtenus à partir des données du registre de l'agglomération de Lille, on observe que les taux obtenus à partir des trois sources étaient équivalents. En revanche, les taux obtenus à partir des données des sources croisées (ALD/ACP)+ et (PMSI/ACP)+ du registre de Lille étaient meilleurs que ceux du Nord-Pas-de-Calais de notre étude (figure 6). L'écart était plus accentué (+14 points) entre les données (ALD/ACP)+ de Lille et notre étude. Ceci pourrait s'expliquer par la présence, en plus des régimes général et indépendant, des données des régimes d'assurance maladie de la MSA et des mines dans les données ALD30 de l'agglomération de Lille.

Figure 7 : Comparaison des taux d'appariement de la région du Nord-Pas-de-Calais avec ceux du registre de l'agglomération de Lille de 2005



1.6 Analyse comparative des données présentes ou absentes des bases croisées

1.6.1 Facteurs favorisant l'enregistrement des cas dans le PMSI

En analyse bivariée, la région de domiciliation, le type histologique, la nature de la pièce opératoire et la taille de la tumeur étaient significativement associés à la présence des cas dans la base PMSI.
En multivariée, les cas de cancer dont la taille des lésions est supérieure à 2 cm ou dépassant la glande étaient 2,3 fois plus présentes dans le PMSI que les cas de taille limitée à la glande. Les cas du Nord-Pas-de-Calais étaient plus présents dans le PMSI qu'en Ile-de-France (tableau 17).

Tableau 17 : Caractéristiques des cas absents du PMSI

Variables	Cas présents dans le PMSI n=343 (%)	Cas absents du PMSI n=80 (%)	Analyse bivariée OR	Analyse bivariée Valeur de p	Analyse multivariée OR	Analyse multivariée IC95 %
Région				**0,001**		
Nord-Pas-de-Calais	153 (88)	20 (12)	1		1	
Ile-de-France	190 (76)	60 (24)	0,4		0,4	[0,2-0,7]
Type histologique				**0,02**		
Papillaire	291 (83)	58 (17)	1			
Vésiculaire	30 (75)	10 (25)	1,7			
Médullaire	13 (72)	5 (18)	0,9			
Autre[a]	9 (56)	7 (44)	0,4			
Nature de la pièce opératoire				**0,008**		
Lobectomie, loboisthmectomie, isthmectomie	45 (83)	9 (17)	1			
Thyroidectomie totale	227 (80)	55 (20)	0,8			
Lobectomie partielle/tumorectomie/ thyroïdectomie sub totale	55 (92)	5 (8)	2,2			
Biopsie thyroïdienne, autre	15 (60)	10 (40)	0,3			
pT				**0,03**		
Taille de la tumeur limitée à la glande (≤ 2 cm)	221 (80)	55 (20)	1		1	
Taille > 2 cm ou dépassant la glande	106 (89)	13 (11)	2,0		2,3	[1,2-4,5]

[a]carcinome mixte, anaplasique, sarcome, SAI, lymphome.

1.6.2 Facteurs favorisant la « mise en ALD » des cas

La région de résidence, le nombre de foyer de lésions et la taille de la tumeur étaient significativement associés à la mise en ALD en analyse bivariée, tandis qu'en multivariée, seule la région de résidence y était significativement associée. Les cas étaient deux fois plus déclarés en ALD en Ile-de-France que dans le Nord-Pas-de-Calais (tableau 18).

Tableau 18 : Caractéristiques des cas absents des ALD

Variables	Cas présents dans ALD n=202 (%)	Cas absents dans ALD n=221 (%)	Analyse bivariée		Analyse multivariée	
			OR	Valeur de p	OR	IC95 %
Région				**0,007**		
Nord-Pas-de-Calais	69 (40)	104 (60)	1		1	
Ile-de-France	133 (53)	117 (47)	1,7		2,0	[1,3-3,1]
Type histologique				**0,09**		
Papillaire	173 (50)	176 (50)	1			
Vésiculaire	17 (42)	23 (58)	1,3			
Médullaire	9 (50)	9 (50)	1,3			
Autre[a]	3 (19)	13 (81)	0,3			
Nombre de foyer de lésions				**0,04**		
Unifocal	128 (45)	157 (55)	1			
Multifocal	73 (56)	58 (44)	1,5			
pT				**0,05**		
Taille de la tumeur limitée à la glande (≤ 2 cm)	124 (45)	152 (55)	1			
Taille > 2 cm ou dépassant la glande	66 (55)	53 (45)	1,5			
Geste ganglionnaire				**0,07**		
Curage	62 (53)	54 (47)	1		1	
Picking	32 (55)	26 (45)	1,1		1,7	[0,8-3,4]
Aucun	101 (43)	136 (57)	0,6		0,7	[0,4-1,2]
Taille de la lésion en mm				**0,07**		
≤ 4	44 (36)	78 (64)	1			
4,1 à 10	58 (54)	49 (46)	2,1			
10,1 à 20	44 (49)	46 (51)	1,7			
> 20	53 (58)	38 (42)	2,5			

[a] carcinome mixte, anaplasique, sarcome, SAI, lymphome.

1.6.3 Caractéristiques des cas ACP absents de la base PMSI et de la base ALD30 (n=69), des cas ACP absents de la base PMSI mais présents dans la base ALD30 (n=11) et des cas ACP absents de la base ALD30 mais présents dans la base PMSI (n=152) (tableau 19)

La distribution des caractéristiques des onze patients exonérés du ticket modérateur absents de la base PMSI suivait celle observée dans la base complète d'analyse constituée, à l'exception des informations sur les caractéristiques relatives à l'encapsulement de la lésion et sa taille. En effet, la majorité des lésions de ces patients avaient une capsule (60 %) et présentait une taille comprise entre 4 et 10 mm (45 %) contrairement aux résultats de l'ensemble des données de la base d'analyse où l'absence de capsule (67 %) et la taille inférieure à 4 mm (30 %) étaient majoritaires.
En ce qui concerne les 152 cas ACP exclusivement présents dans la base PMSI, les 69 cas ACP absents des bases PMSI etALD30 et les 191 cas présents à la fois dans les ALD et PMSI, la distribution suivait celle des données de la base d'analyse à l'exception des informations sur la taille des lésions des 191 cas présents à la fois dans les ALD et PMSI (ACP/ALD/PMSI)+. En effet les tailles comprises entre 4 et 10 mm ou supérieures à 20 mm étaient majoritaires parmi les cas (ACP/ALD/PMSI)+ comparées aux cas ACP pour lesquels la taille ≤ 4 mm était majoritaire.

Tableau 19 : Caractéristiques des cas ACP présents dans la base ALD30 ou PMSI et absents de la base ALD30 et de la base PMSI

	(ALD/ACP)+PMSI- N = 11 (%)	(PMSI/ACP)+ALD- N = 152 (%)	ACP+ALD-PMSI- N = 69 (%)	(ACP/ALD/PMSI)+ N = 191 (%)	ACP de référence N = 423 (%)
Sexe					
Masculin	1 (9)	39 (26)	19 (28)	40 (21)	99 (23)
Féminin	10 (91)	113 (74)	50 (72)	151 (79)	324 (77)
Âge					
< 45 ans	4 (36)	51 (34)	24 (35)	79 (41)	158 (37)
≥ 45 ans	7 (63)	101 (66)	45 (65)	112 (59)	265 (63)
Région					
Nord-Pas-de-Calais	1 (9)	67 (44)	19 (28)	68 (36)	173 (41)
Ile-de-France	10 (91)	85 (56)	50 (72)	123 (64)	250 (59)
Type histologique					
Vésiculaire	-	13 (8)	10 (15)	17 (9)	40 (9)
Papillaire	11 (100)	129 (85)	47 (69)	162 (85)	349 (83)
Médullaire	-	4 (3)	5 (7)	9 (5)	18 (4)
Autre	-	6 (4)	6 (9)	3 (1)	16 (4)
Métastase					
Pas de métastase	2 (25)	12 (9)	6 (11)	22 (14)	42 (12)
Renseignements insuffisants	6 (75)	122 (91)	48 (89)	140 (86)	316 (88)
Poids de la lésion en gramme					
≤ 20	2 (22)	47 (33)	15 (25)	67 (38)	131 (34)
Entre 20,1 et 50	7 (78)	64 (45)	30 (51)	79 (44)	180 (46)
> 50	-	32 (22)	14 (24)	32 (18)	78 (20)
Nombre de foyer de lésions					
Unifocal	8 (73)	113 (75)	44 (69)	120 (63)	285 (69)
Multifocal	3 (27)	38 (25)	20 (31)	70 (37)	131 (31)
Lésion encapsulée					
Oui	6 (60)	44 (30)	17 (28)	65 (35)	132 (33)
Non	4 (40)	103 (70)	43 (72)	123 (65)	273 (67)
Invasions vasculaires					
Non	7 (87)	58 (74)	37 (80)	114 (81)	216 (79)
Oui douteux	-	3 (4)	1 (2)	4 (3)	8 (3)
Oui manifeste	1 (13)	17 (22)	8 (18)	23 (16)	49 (18)
Invasion extra thyroïdienne					
Non	9 (100)	128 (85)	51 (86)	149 (82)	337 (84)
Oui	-	22 (15)	8 (14)	33 (18)	63 (16)
Taille de la lésion en mm					
≤ 4	4 (37)	56 (37)	22 (37)	40 (21)	122 (30)
4,1 à 10	5 (45)	32 (21)	17 (28)	53 (28)	107 (26)
10,1 à 20	2 (18)	35 (23)	11 (18)	42 (23)	90 (22)
> 20	-	28 (19)	10 (17)	53 (28)	91 (22)

1.7 Résultats de l'exercice de retour au cas

Sur les huit structures ACP du pilote sollicitées pour l'exercice de retour au cas, seules 6 (4 en IDF et 2 dans le NPdC) avaient favorablement répondu.

Sur les 100 cas tirés, un a été exclu de l'exercice car son numéro de fiche ne permettait pas d'identifier clairement la structure ACP à laquelle il appartenait. La totalité des cas transmis aux structures ayant participé à l'exercice a été retrouvée dans les dossiers de chaque structure ACP répondante (tableau 20).

Tableau 20 : Répartition des cas de cancer de la thyroïde tiré au sort par structure ACP et résultats de l'exercice de retour au cas

Structure ACP	Effectif tiré au sort et transmis aux ACP (%)	Structure ACP ayant répondu	% effectif retrouvé
Ile-de-France	**59 (60)**		
AP-HP Pitié Salpétrière	21 (21)	Oui	100
AP-HP Ambroise Paré	5 (5)	Oui	100
Institut Gustave Roussy	19 (19)	Oui	100
Laboratoire Tolbiac	12 (12)	Oui	100
Institut de pathologie de Paris	2 (2)	Non	-
Nord-Pas-de-Calais	**40 (40)**		
Centre hospitalier régional de Lille	38 (38)	Oui	100
Laboratoire de Valenciennes	1 (1)	Non	-
GH de l'Institut catholique de Lille	1 (1)	Oui	100
Total	**99 (100)**		

2 Transmission des fiches ACP de l'année 2009 pour le volet prospectif du pilote

Le volet prospectif du pilote n'avait aucun objectif de croisement. Il permettait de tester en routine l'anonymisation et la remontée sécurisée des données ACP via l'application informatique installée.
Sur les 8 structures pilotes seules 6 (Pitié Salpétrière, Ambroise Paré, Institut Gustave Roussy, Centre hospitalier régional et universitaire de Lille, Laboratoire de Valenciennes, Groupe hospitalier de l'Institut catholique de Lille) ont participé au volet prospectif portant sur le recueil des données entre le 1er février et le 1er mai 2009. Durant cette période 74 cas de cancer de la thyroïde ont été enregistrés, anonymisés et transmis d'une manière sécurisée à l'InVS. Aucune anomalie fonctionnelle d'émission ou de réception n'a été signalée.

3 Discussion

La localisation cancéreuse (thyroïde) étudiée dans le pilote SMSC ne constitue pas un problème prioritaire de santé publique à cause d'une part de sa faible fréquence (6 672 nouveaux cas estimés en 2005) [14] et d'autre part de sa faible létalité (403 décès en 2005) [14]. La pertinence de ce choix se justifiait par le fait que son diagnostic est toujours posé à l'occasion d'un examen anatomo pathologique et que le traitement relève systématiquement de la chirurgie. De plus, sa faible fréquence est à l'origine d'un faible volume d'enregistrement qui génère moins de charge de travail aux pathologistes impliqués dans le pilote.

L'objectif du SMSC est la surveillance épidémiologique des cancers par le dénombrement des nouveaux cas de cancers en tout point du territoire français. Le croisement des trois sources de données nécessitait de connaître le niveau de cohérence des données croisées et le pourcentage des données communes entre ces sources.
L'analyse du niveau de cohérence des données a reposé dans notre étude sur les informations administratives telles que le sexe, l'âge et le lieu de résidence du cas.
Le pourcentage de cohérence sur le sexe et le département de résidence entre les trois sources de données était très élevé, il variait de 97 % à 100 %. Les discordances, certes marginales, observées pourraient s'expliquer soit par des erreurs de saisie soit par la prise en charge hospitalière d'un bénéficiaire différent de l'ayant droit.
L'âge des données ALD et ACP était calculé en « âge atteint » dans l'année alors que celui du PMSI était renseigné en « âge révolu » surestimant d'un an les données ALD et ACP. La prise en compte de cette surestimation montre une cohérence parfaite de l'âge entre les trois sources.

Le taux d'appariement des données croisées s'élevait à 84% mais n'était pas identique dans les deux régions étudiées à cause probablement de différences de pratiques médicales dans ces deux régions. L'absence de certains cas dans les bases médico administratives analysées pourrait s'expliquer par plusieurs facteurs liés au mode de prise en charge hospitalière de la pathologie étudiée ou à l'exonération du ticket modérateur pour affection longue durée.

3.1 La prise en charge hospitalière

- La porosité des frontières interrégionales : notre étude s'est déroulée dans deux régions pilotes, l'hypothèse selon laquelle le diagnostic ACP est posé dans une région et les soins administrés hors de la région ne pourrait être écartée.

En effet, un patient dont le lieu de résidence est proche d'une deuxième région administrative peut se faire diagnostiquer dans la première et bénéficier des soins dans la deuxième où l'offre de soins spécifique à la maladie est considérée comme plus performante. Cette mobilité interrégionale des soins a pour conséquence l'absence dans la région de résidence des données d'hospitalisation du patient traité. Cependant cet écueil disparaîtrait lorsque le traitement des données sera réalisé au niveau national.

- Les données manquantes : dans notre étude, 4 % des identifiants patients de la base PMSI sont invalides et supprimés de la base avant croisement. L'on ne pourrait pas exclure la possibilité d'appariement entre ces séjours et les cas ou les données ALD s'ils avaient été valides, ce qui augmenterait le taux d'appariement du croisement.
- Les hospitalisations de fin d'année : certains patients diagnostiqués pour cancer de la thyroïde sont effectivement hospitalisés pour une prise en charge dont la période se situe à cheval entre deux années. Le PMSI ne prenant en compte que les périodes de sortie des patients, il y a un risque que les cas dont la date de découverte se situe à l'année n soient comptabilisés dans les données de l'année suivante (n+1). Par ailleurs, certains types de cancers de « très bon pronostic » peuvent faire l'objet d'une prise en charge différée entre le moment du diagnostic et le traitement. Lorsque le délai entre la période de diagnostic et la période du traitement chirurgical est long et se situe à cheval entre deux années, les informations sur le patient seront absentes des données hospitalières de l'année du diagnostic.
- Décalage entre le codage du séjour et le diagnostic histologique du cancer : les résultats histologiques de la lésion traitée peuvent arriver après le codage du séjour hospitalier, codage qui pourrait ne pas faire mention de la nature maligne de la tumeur empêchant son repérage parmi les données de cancer de la thyroïde du PMSI.

3.2 La mise en ALD

La mise en ALD n'est pas obligatoire, de ce fait certains assurés peuvent volontairement choisir de ne pas en bénéficier donc ne pas apparaître dans la base des données des ALD. D'autres en sont absents pour cause d'absence de couverture sociale (très exceptionnelle) ou de déclaration de la maladie à une caisse d'assurance non domiciliée à leur lieu de résidence. Pour d'autres la demande n'a pas été faite car le traitement radical ne nécessitait pas de suivi thérapeutique coûteux.
Par ailleurs, les cas diagnostiqués en fin d'année par les ACP ont de forte chance de ne pas figurer dans la base ALD30 de la même année si l'avis d'exonération du ticket modérateur pour affection longue durée est rendu en début de l'année suivante.
Les données extraites du Sniiram dans le cadre de ce projet proviennent des régimes général et indépendant et ne comprennent pas les données individuelles du régime agricole. Or l'effectif des données agrégées 2004 du régime agricole dans les deux régions d'étude (disponible à l'InVS) s'élevait à 16 cas de cancer thyroïdien. Dans l'hypothèse où la totalité de ces cas serait présente dans l'une des bases ACP ou PMSI, elle augmenterait le taux global d'appariement de 3 points c'est-à-dire 87 %.

En comparant les taux d'appariement de la région Nord-Pas-de-Calais de notre étude à ceux du registre de l'agglomération de Lille de l'année 2005 dont les données ALD proviennent de quatre régimes à savoir l'échelon régional du régime général, le RSI, la MSA et le régime spécial des mines, on observe deux situations : 1) les taux d'appariement entre les trois sources de données sont très proches (89 % *versus* 88 %), 2) les taux d'appariement entre les sources croisées deux à deux impliquant la source ACP sont plus élevés dans le registre de l'agglomération de Lille que dans notre étude avec un différentiel de 8 points (PMSI/ACP)+ et de 14 points (ALD/ACP)+. Le plus grand écart observé concerne les cas croisés avec les données ALD dont les régimes de provenance tels que la MSA et les mines sont absents de notre étude. Ceci souligne la sensibilité du taux d'appariement vis-à-vis des régimes de données ALD dont les spécificités régionale et départementale doivent être prises en compte dans l'estimation infra nationale de l'incidence des cancers à partir du système automatisé (SMSC) en cours de développement.

V - CONCLUSION ET PERSPECTIVES

1 Malgré les difficultés, le pilote a atteint ses objectifs d'étude de faisabilité

1.1 Pertinence

La réalisation de cette étude a permis de répondre aux objectifs initiaux en termes de pertinence et de faisabilité, et notamment de :

- mettre en place une solution informatique complète pour la remontée des données ACP anonymisées, pour constituer une base de données test à l'InVS ;
- obtenir les données du PMSI et du Sniiram avec le numéro d'anonymat commun et évaluer la qualité de l'appariement des trois sources ;
- tester la faisabilité du retour au cas ;
- mettre en évidence les points critiques au déploiement pour envisager les modalités les plus pertinentes pour la poursuite des travaux vers un système de configuration nationale.

La mise en place du croisement des trois sources anonymisées au niveau national répondait à un besoin à la fois scientifique et institutionnel fort ; l'étude a démontré que, pour la localisation testée, le cancer thyroïdien, le croisement permettait d'obtenir des résultats en termes de repérage des cas proches de ceux des registres généraux.
Le travail sur la stratégie de sélection des cas incidents reste à faire ; le développement d'algorithmes de sélection des cas incidents est en cours par le DMCT pour les bases PMSI et ALD.

1.2 Faisabilité

En termes de faisabilité, si l'étude a démontré qu'un tel dispositif pouvait être opérationnel à une échelle limitée (huit structures ACP, deux régions), elle a cependant mis en évidence que la solution testée atteignait un niveau de complexité tel qu'il n'était pas envisageable de la déployer telle quelle au niveau national. La durée de l'étude (5 ans et demi, depuis les travaux préparatoires jusqu'à la fin de la prestation informatique) rend compte pour partie de cette complexité à la fois technique et retrouvée à tous les autres niveaux, qu'ils soient conceptuel, juridique, contextuel ou politique.

1.2.1 Au niveau conceptuel

La première difficulté trouvait son origine dans le principe même du projet, à savoir le croisement de bases de données par l'utilisation d'un identifiant universel de santé pour permettre un appariement sécurisé, cet identifiant n'existant pas en routine.
Cette problématique est particulièrement prégnante pour l'InVS, qui, pour répondre à ses missions, a besoin de pouvoir croiser entre elles des informations individuelles anonymisées issues de bases médico administratives existantes, ou avec d'autres sources qu'il met en place avec ses partenaires du réseau national de santé publique, afin de produire les indicateurs nécessaires à l'orientation et au pilotage des politiques publiques.
Les difficultés rencontrées dans le cadre de l'étude pilote, qui dépassent largement le cadre du projet SMSC, étaient notamment liées aux contraintes de la Cnil dans l'accessibilité et l'utilisation du NIR ou d'identifiants (comme Foin) basés sur le NIR pour le croisement de données médicales personnelles. Des contacts avec la Cnil à ce sujet ont amené le président de la Cnil, M. Alex Türk, à proposer à l'InVS d'exposer les difficultés rencontrées dans l'exercice de ses missions en séance plénière de la Commission. Cette audition, première étape d'une réflexion de fond sur ce sujet, a eu lieu le 15 octobre 2009. Depuis, des textes sont en préparation pour permettre l'accès de l'InVS au NIR ainsi qu'aux données individuelles de l'Assurance maladie, dans le cadre de ses missions.

1.2.2 Au niveau technique et juridique

La lourdeur du cadre contractuel, avec 25 documents de tous types encadrant une étude de faisabilité au périmètre limité, était sous-tendue par la forte dimension multi partenariale du projet, les partenaires intervenant à tous les niveaux du dispositif : participation au circuit de remontée des données, cession des outils techniques et des données, intégration du SMSC sur des réseaux protégés…

Les principales difficultés techniques peuvent être résumées ci-dessous :

- les pré-requis n'étaient pas disponibles au départ, que ce soit les modules Foin et Camel (avec par ailleurs impossibilité pour la société prestataire W d'avoir accès à Camel et donc de pouvoir l'installer en environnement de test) ou la documentation (initialement absente puis constatation que la version communiquée était obsolète) ;
- la solution développée était à la carte en fonction des sites, non totalement automatisée, avec des connecteurs pour certains semi-manuels voire manuels ;
- les flux en particulier étaient instables, rompant toute la chaîne de transmission ;

- le manque de maîtrise des SI des partenaires (parfois soumis à une infogérance, à des changements de SIH en cours d'étude, des réorganisations de service, des modifications de paramétrage, changement de parc informatique, débordement des services informatiques des établissements…) a conduit à une maintenance difficile de la solution déployée ;
- les fortes contraintes techniques autour des outils Foin et Camel, de l'intégration du SMSC dans des SI très sécurisés, ont compliqué l'intégration d'environnements informatiques hétérogènes, en l'absence de possibilité d'utiliser largement les technologies libres basées sur le web ;
- la coordination technique et la synchronisation des développements à réaliser par les partenaires et en lien avec l'avancée de la prestation informatique ont été particulièrement délicates et ont largement impacté le déroulement de la prestation et ont conduit *in fine* à un retard de 15 mois par rapport au planning initial.

1.2.3 Au niveau contextuel et politique

Le SMSC suppose une structuration des données ACP pour constituer une base exploitable sur le plan informatique et statistique. Cette structuration nécessitait :

- d'avoir défini les items pertinents à renseigner, à la fois pour le SMSC mais bien entendu avant tout pour l'utilisation du compte-rendu en pratique clinique ;
- d'intégrer dans les logiciels métiers ACP les items et le module d'extraction ;
- d'implanter les versions actualisées des logiciels métiers chez les ACP.

Par ailleurs, si le développement du SMSC était une des premières mesures du plan cancer 2003-2007, le plan prévoyait, en parallèle, le développement d'autres projets basés sur l'ACP, axés sur l'amélioration de la prise en charge des patients et pilotés par l'INCa : les projets CRFS et DCC. Sans être incompatibles, les deux projets devaient nécessairement s'articuler très étroitement, puisqu'ils avaient les mêmes acteurs, les médecins ACP, et le même moyen de transmission, leurs logiciels métiers. Alors que le SMSC visait un test sur une localisation avec un développement rapide des aspects informatiques, le projet DCC travaillait d'emblée sur l'ensemble des localisations cancéreuses avec la profession ACP.
Ainsi, les projets SMSC, DCC et CRFS, tous issus de la volonté gouvernementale et inscrits dans le plan cancer ont débuté ensemble, mais avec des objectifs opérationnels, des méthodologies et des calendriers très distincts. L'indispensable articulation entre ces projets a été d'autant plus difficile à organiser qu'elle intervenait dans un contexte de revendication par la profession des ACP d'une meilleure reconnaissance de leur activité, faisant face à une baisse des effectifs des professionnels et à un retard dans la mise en place de la classification commune des actes médicaux pour la spécialité (CCAM-ACP).

2 Une évolution du contexte autour des systèmes d'information partagée

L'étude pilote, débutée fin 2004, s'est achevée début 2010. En cinq ans, le contexte politique autour des systèmes d'information partagée a été fortement restructuré, avec en particulier la naissance de l'Agence des systèmes d'information partagée en santé (ASIP) en septembre 2009 reprenant le chantier de la mise en place du DMP dans les suites du GIP-DMP. Sous l'impulsion du plan cancer 2009-2013 (mesure 18.3 « Partager les données médicales entre professionnels de santé ») et de la relance du DMP, la maitrise d'ouvrage de la mise en place du DCC a été confiée par l'INCa à l'Asip Santé dans un accord de partenariat signé en décembre 2009.
A l'issue d'un travail de concertation entre les équipes de l'INCa et de l'ASIP Santé, mais également avec les réseaux régionaux de cancérologie, chargés de l'organisation territoriale de l'accompagnement des patients atteints de cancer, **un cadre national** a été publié par les deux organismes en octobre 2010. Le DCC est désormais conçu comme un service spécialisé du DMP, au service de la coordination des soins pour les patients atteints de cancer. Il établit un lien privilégié entre chaque professionnel de santé, qu'il relève de la médecine de ville comme de l'hôpital, afin d'accompagner la prise en charge du patient pendant et après la phase aiguë de son traitement.
Le programme de travail DCC/DMP inclut la mise en œuvre d'une phase pilote pour vérifier et finaliser le fonctionnement organisationnel et technique de ce service, en capitalisant notamment sur les travaux et acquis des Réseaux régionaux de cancérologie (RRC) et pour appuyer le démarrage du DMP.
sept régions pilotes (constituées des RRC et de la maîtrise d'ouvrage régionale des systèmes d'information) soutenues par leurs ARS ont été retenues : Alsace, Aquitaine, Lorraine, Midi-Pyrénées, Pays-de-la-Loire, Picardie et Rhône-Alpes, et devront mener leurs travaux sur une période de 18 à 24 mois.

2.1 Standardisation des comptes-rendus ACP

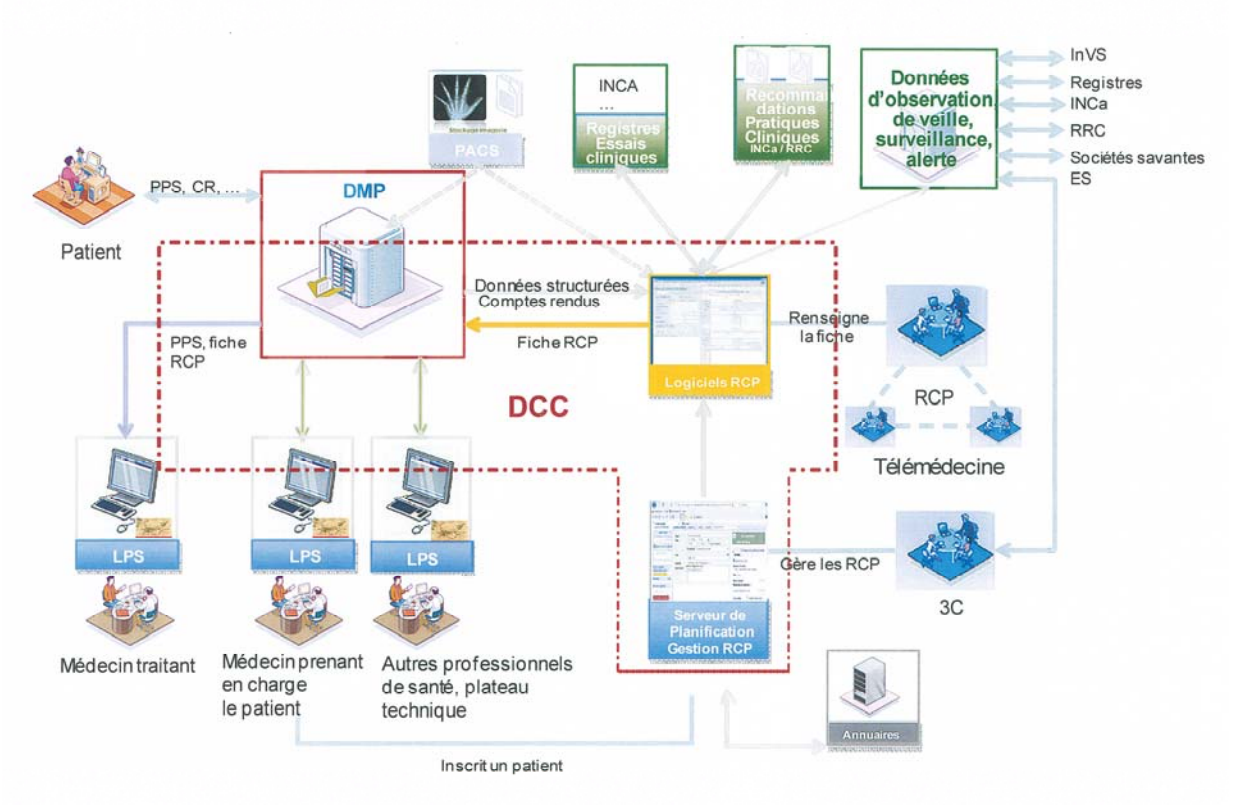
L'INCa mène depuis plusieurs années avec la profession des ACP des travaux de standardisation des comptes-rendus ACP, qui doivent alimenter le DCC et constituent la principale des sources du SMSC. L'InVS y a été associé pour la définition des items indispensables à la surveillance épidémiologique et à l'identification d'un patient dans le SMSC.
Récemment ces travaux ont abouti, dans le cadre du dispositif d'autorisation des établissements de santé pour le traitement du cancer, à l'élaboration, par la Société française de pathologie, de listes d'items minimaux par localisation

cancéreuse[11]. Ces fiches ont été regroupées en décembre 2009 sous forme d'un catalogue (« Comptes-rendus ACP : données minimales à renseigner pour une tumeur primitive »), mis en ligne sur le site de l'INCa [15].

2.2 Structuration des comptes-rendus ACP

L'ASIP Santé travaille à la mise en place d'un répertoire national des référentiels, dont l'ACP ; elle a compétence pour porter la structuration des données des comptes-rendus ACP, définies par l'INCa et la SFP. Ces données structurées seront à terme intégrées dans les logiciels métiers des ACP afin d'alimenter le DCC.
La figure 7 montre l'architecture des flux prévue dans le cadre du DCC. Il est prévu la création d'entrepôts de données alimentés à partir des réunions de concertation pluridisciplinaires (RCP), qui ont lieu dans les établissements de soins. Les données structurées ACP, transmises au logiciel RCP à partir du DCC seront extraites de ces entrepôts pour constituer une base nationale structurée, qui sera accessible à l'InVS pour le SMSC, sous une forme anonymisée, mais aussi aux registres et aux structures de gestion des dépistages (sous forme nominative).
Réglementairement, les données ACP en cours de structuration ne comporteront pas le NIR comme numéro d'anonymat mais l'INS (identifiant national de Santé) calculé ou aléatoire. Cet identifiant ne permettra pas de croiser les bases ALD/PMSI et ACP. Toutefois, si la base de données structurée accessible à l'InVS pour l'exercice de ses missions est constituée à partir des RCP, situés dans les établissements de soins, il devrait être possible de récupérer dans le SIH de l'établissement le numéro d'anonymat anoSNIIR1, permettant, lui, le croisement des données ACP avec le PMSI et les ALD.

Figure 8 : Architecture des flux DCC, étude pilote SMSC, 2004-2010 (source : INCa/Asip santé, DCC et DMP, cadre national, 2010)



3 Les perspectives

Si le dispositif SMSC cible reste inchangée, avec trois sources de données croisées, la poursuite des travaux va toutefois comporter deux étapes distinctes, à développer en parallèle, la première cherchant à produire des indicateurs de surveillance des cancers à partir des sources ALD et PMSI, la seconde visant à mettre en place la source ACP, en lien avec le déploiement du DCC.

[11] Ces listes d'items ont été établies pour 21 localisations tumorales, représentant 85 % des nouveaux cas de cancer

3.1 1er axe de développement : croisement des deux sources PMSI et ALD au niveau national

Dans l'immédiat, le premier axe sera de poursuivre les travaux autour du croisement des deux sources médico administratives déjà disponibles (PMSI et ALD) à des fins descriptives et de production d'indicateurs de surveillance des cancers.

3.2 2e axe de développement : mise en place du SMSC trois sources

Le deuxième axe, à moyen terme et en parallèle avec l'axe 1 étudiera les conditions de mise en place d'une base nationale de données ACP, en vue de son croisement avec les deux sources précédentes. Le seul dispositif viable envisageable à l'heure actuelle pour constituer la source ACP sera un système utilisant des flux existants dans le cadre du DCC. Le recueil de l'adresse des patients en vue d'un géocodage permettrait de répondre aux questions d'exposition environnementale mais aussi d'inégalités sociales. Un des enjeux principaux sera d'utiliser un identifiant national qui soit commun aux bases PMSI et ALD30 afin de permettre le croisement des données. Ce pourrait être un identifiant issu de Foin.

Le SMSC constitue un système de surveillance de troisième génération, les systèmes multi sources. La problématique du croisement de bases de données médico-administratives entre elles, avec d'autres sources de données créées de novo ou d'études de cohortes est une problématique émergente. Avec elle, se pose la question d'un identifiant national qui permettent le rapprochement de ces sources. La « communication » entre eux des différents systèmes d'information en santé existants ou à construire, dans les limites du respect de la confidentialité des données, représente un des défis majeurs des années à venir en matière de surveillance et de recherche en santé publique.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] Hospices civils de Lyon/Institut de veille sanitaire/Institut national du cancer/Francim/Institut national de la santé et de la recherche médicale. Projections de l'incidence et de la mortalité par cancer en France en 2010. Rapport technique. 2010. [consulté le 19/08/2010].

[2] Belot A, Grosclaude P, Bossard N, Jougla E, Benhamou E, Delafosse P, et al. Cancer incidence and mortality in France over the period 1980-2005. Rev Epidemiol Santé Publique 2008;56(3):159-175.

[3] Le plan cancer 2009-2013. Institut national du cancer, France [mis à jour le 2009; consulté le 19/08/2010].
Disponible à partir de l'URL : http://www.e-cancer.fr/plancancer-2009-2013

[4] Leenhardt L, Grosclaude P, Chérié-Challine L, et al. Recommandations pour la mise en place d'un dispositif de surveillance épidémiologique national des cancers thyroïdiens. Institut de veille sanitaire, France [mis à jour le 01/04/2003; consulté le 17/06/2011].
Disponible à partir de l'URL : http://www.invs.sante.fr

[5] Leenhardt L, Grosclaude P, Chérié-Challine L, et al. Mise en place d'un dispositif de surveillance épidémiologique nationale des cancers thyroïdiens. Institut de veille sanitaire, France [mis à jour le 2001; consulté le 08/2010].
Disponible à partir de l'URL : http://www.invs.sante.fr

[6] Verger P, Chérié-Challine L, et al. Cancer de la thyroïde en France et accident de Tchernobyl: évaluation des risques potentiels et recommandations pour le renforcement des connaissances épidémiologiques. Bull Epidemiol Hebd 2001;2197-99.

[7] Verger P, Chérié-Challine L. Evaluation des conséquences sanitaires de l'accident de Tchernobyl en France. Dispositif de surveillance épidémiologique, état des connaissances, évaluation des risques et perspectives. Institut de veille sanitaire, France [mis à jour le 01/01/2000; consulté le 04/07/2011].
Disponible à partir de l'URL : http://www.invs.sante.fr

[8] Structure du fichier des tumeurs, sources de données. Registre des tumeurs du Québec [mis à jour le 2010; consulté le 12/10/2010].
Disponible à partir de l'URL :
http://msssa4.msss.gouv.qc.ca/santpub/tumeurs.nsf/61a4a0842e5cbd34852568d500653357/bea4e41a3066f163852568d900660b4b?OpenDocument

[9] Geoffroy-Perez B. Confrontation des données du programme national de surveillance du mésothéliome et du PMSI. Institut de veille sanitaire, France [mis à jour le 01/09/2004; consulté le 04/07/2011].
Disponible à partir de l'URL : http://www.invs.sante.fr

[10] Carré N, Uhry Z, Velten M, Trétarre B, Schvartz C, Molinié F, et al. Valeur prédictive et sensibilité du programme de médicalisation des systèmes d'information (PMSI) par rapport aux registres des cancers : application au cancer de la thyroïde (1999-2000). Rev Epidemiol Santé Publique 2006;367-376.

[11] Grosclaude P, et al. Evaluation de la qualité des ALD comme indicateurs de suivi épidémiologique des cancers à l'aide de données nominatives. Rapport. 2010. [consulté le 06/2010].

[12] Grosclaude P, et al. Evaluation de la qualité du PMSI comme indicateurs de suivi épidémiologique des cancers à l'aide de données nominatives. Rapport. 2010. [consulté le 07/2010].

[13] Uhry Z, Colonna M, Remontet L, Grosclaude P, Carré N, Couris CM, et al. Estimating infra national and national thyroid incidence in France from cancer registries data and national hospital discharge database. Eur J Epidemiol 2007;22607-614.

[14] Cancer de la thyroïde, estimation en 2005, principaux indicateurs. Institut de veille sanitaire, France [mis à jour le 2010; consulté le 12/10/2010].
Disponible à partir de l'URL :
http://www.invs.sante.fr/surveillance/cancers/estimations_cancers/default.htm

[15] Comptes-rendus d'anatomopathologie: données minimales à renseigner pour une tumeur primitive, traitement, soins et innovations. Institut national du cancer, France [mis à jour le 2009; consulté le 19/08/2010].
Disponible à partir de l'URL : http://www.e-cancer.fr

ANNEXES

ANNEXE 1 – FICHE ÉPIDÉMIOLOGIQUE THYROÏDE, STRUCTURES ACP PRIVÉES

Surveillance épidémiologique nationale des cancers thyroïdiens

Items ACP à extraire pour chaque nouveau cas histologique de cancer thyroïdien – Structure ACP privée

PATIENT

SEXE /___/ M (1), F (2),
DATE DE NAISSANCE /__ __//__ __//__ __ __ __/
N°SS /__//__ __//__ __//__ __//__ __ __//__ __ __/ /__ __/
COMMUNE DE DOMICILE en clair CODE POSTAL
/_____/ /_____/

IDENTIFICATION FICHE

N° FINESS STRUCTURE N°FICHE
/_ _ _ _ _ _ _ _ _/ /_ _ _ _ _ _ _ _ _ _/

NOM DU MEDECIN AYANT REMPLI LA FICHE : /_____/
DATE DE SAISIE DE LA FICHE : /__ __//__ __//__ __ __ __/

Nature de la pièce opératoire
1 : lobectomie / loboisthmectomie
2 : thyroïdectomie totale
3 : isthmectomie
4 : lobectomie partielle / tumorectomie
5 : thyroïdectomie subtotale
6 : biopsie thyroïdienne
7 : autre

1 - LE CANCER THYROÏDIEN

1.1 Date de l'enregistrement du prélèvement /__ __//__ __//__ __ __ __/

1.2 Nature de la pièce opératoire (voir encadré ci-contre) /___/

1.3 Poids de la pièce opératoire (g) /__ __ __/ et/ou **Taille (mm)** LD /__ __ __/ LG /__ __ __/

1.4 Nombre total de prélèvements sur la thyroïde examinée (ou nombre de blocs si 1 prélèvement/bloc) /__ __/ (si NP : 99)

1.5 Nombre total de prélèvements sur la lésion /__ __/ (si NP : 99)

1.6 Geste ganglionnaire /___/ curage (1), picking (2), aucun (0), NP (9)

2 - LA LESION

2.1 Taille de la lésion la plus grande (mm) /__ __ __/ **mm**

2.2 Focalité /____/ unifocale (1), multifocale (2), multifocale avec variante morphologique (3)

2.3 Localisation **droite** non(0), oui(1), NP(9) /____/ **gauche** non(0), oui(1),NP(9) /____/ **médiane** non(0), oui(1), NP(9) /____/

2.4 Capsule de la lésion /____/ non encapsulée (0), encapsulée (1), NP (9)
Si encapsulée /____/ non infiltrée (0), infiltrée non dépassée (1), dépassée (2), NP (9)

2.5 Invasions vasculaires /____/ non (0), douteux (1), oui manifeste (2), NP (9)
Si oui manifeste, précisez si ≥ 4 /____/ non (0), oui (1), NP (9)

2.6 Invasions extra thyroïdiennes /____/ non (0), oui T3 (1), oui T4 (2), NP (9)

2.7 Si ganglions, précisez **nombre** /__ __/ (si NP : 99) **dont ganglions envahis N+** /__ __/
Si N+, précisez si
gg lymphatiques régionaux homolatéraux /____/ non (0), oui (1), NP (9)
prétrachéaux / paratrachéaux / prélaryngés /____/ non (0), oui (1), NP (9)
autres chaînes cervicales ou médiastinales sup /____/ non (0), oui (1), NP (9)

2.8 pTNM **pT** /____/ **pN** /____/ **M** /____/ (si NP : 9) **numéro de l'édition utilisée** /____/

3 - DIAGNOSTIC FINAL [A] /____/ **merci d'utiliser les n° de la rubrique « diagnostic final »**
2ème avis demandé /____/ non (0), oui (1) **si oui, diagnostic final posé après 2ème avis** /____/ non (0), oui (1)

4 - LES LESIONS ASSOCIEES [B] /____/ /____/ /____/
NB si autre cancer thyroïdien associé (8) remplir : **diagnostic final du cancer associé** /____/

[A] **Diagnostic final** selon OMS et/ou AFIP (tous ces numéros seront transcodés automatiquement en ADICAP, SNOMED ou CIM10)

1.carcinome vésiculaire SAI
2.carcinome vésiculaire à invasion minime, encapsulé
3.carcinome vésiculaire largement invasif
4.carcinome vésiculaire à cellules oxyphiles, éosinophiles, oncocytaire, à cellules de Hürtle
5.carcinome vésiculaire à cellules claires
6. carcinome vésiculaire peu différencié et/ou insulaire
7.carcinome papillaire SAI
8.carcinome papillaire à variante vésiculaire
9.carcinome papillaire sclérosant diffus, occulte, non encapsulé
10.carcinome papillaire à cellules oxyphiles, éosinophiles
11.carcinome papillaire à cellules cylindriques, hautes
12.carcinome papillaire à présentation solide et/ou trabéculaire
13.carcinome médullaire SAI
14.carcinome mixte médullaire vésiculaire
15. carcinome mixte médullaire papillaire
16.carcinome médullaire héréditaire
17.carcinome anaplasique sans contingent bien différencié identifiable
18.carcinome anaplasique avec un contingent bien différencié minoritaire
19.carcinome anaplasique avec un contingent bien différencié majoritaire
20.carcinome indifférencié
21. carcinome SAI
22.lymphome
23.sarcome

[B] **Lésions associées**

1.absence
2.hyperplasie diffuse ou nodulaire
3.thyroïdite lymphocytaire
4.adénome SAI
5.adénome atypique
6.adénome à cellules oxyphiles
7. adénome trabéculaire hyalinisant
8.autre cancer thyroïdien de type histologique différent
9.lésion fonctionnelle (Basedow, hyperplasie diffuse ou nodulaire)
10.nodule toxique

ANNEXE 2 – FICHE ÉPIDÉMIOLOGIQUE THYROÏDE, STRUCTURES ACP PUBLIQUES

Surveillance épidémiologique nationale des cancers thyroïdiens

Items ACP à extraire pour chaque nouveau cas histologique de cancer thyroïdien – Structure ACP publique

PATIENT

SEXE /___/ M (1), F (2),
DATE DE NAISSANCE /__ __//__ __//__ __ __ __/
ANONYMAT* /_____/
COMMUNE DE DOMICILE en clair CODE POSTAL
/_____/ /_____/

IDENTIFICATION FICHE

N° FINESS STRUCTURE N°FICHE
/_ _ _ _ _ _ _ _ _/ /_ _ _ _ _ _ _ _ _ _/

* le numéro d'anonymat est transmis automatiquement par le système informatique de l'hôpital

NOM DU MEDECIN AYANT REMPLI LA FICHE : /_____/
DATE DE SAISIE DE LA FICHE : /__ __//__ __//__ __ __ __/

Nature de la pièce opératoire
1 : lobectomie / loboisthmectomie
2 : thyroïdectomie totale
3 : isthmectomie
4 : lobectomie partielle / tumorectomie
5 : thyroïdectomie subtotale
6 : biopsie thyroïdienne
7 : autre

1 - LE CANCER THYROÏDIEN

1.1 Date de l'enregistrement du prélèvement /__ __//__ __//__ __ __ __/

1.2 Nature de la pièce opératoire (voir encadré ci-contre) /___/

1.3 Poids de la pièce opératoire (g) /__ __ __/ et/ou **Taille (mm)** LD /__ __ __/ LG /__ __ __/

1.4 Nombre total de prélèvements sur la thyroïde examinée (ou nombre de blocs si 1 prélèvement/bloc) /__ __/ (si NP : 99)

1.5 Nombre total de prélèvements sur la lésion /__ __/ (si NP : 99)

1.6 Geste ganglionnaire /___/ curage (1), picking (2), aucun (0), NP (9)

2 - LA LESION

2.1 Taille de la lésion la plus grande (mm) /__ __ __/ **mm**

2.2 Focalité /____/ unifocale (1), multifocale (2), multifocale avec variante morphologique (3)

2.3 Localisation **droite** non(0), oui(1), NP(9) /____/ **gauche** non(0), oui(1), NP(9) /____/ **médiane** non(0), oui(1), NP(9) /____/

2.4 Capsule de la lésion /____/ non encapsulée (0), encapsulée (1), NP (9)
Si encapsulée /____/ non infiltrée (0), infiltrée non dépassée (1), dépassée (2), NP (9)

2.5 Invasions vasculaires /____/ non (0), douteux (1), oui manifeste (2), NP (9)
Si oui manifeste, précisez si ≥ 4 /____/ non (0), oui (1), NP (9)

2.6 Invasions extra thyroïdiennes /____/ non (0), oui T3 (1), oui T4 (2), NP (9)

2.7 Si ganglions, précisez **nombre** /__ __/ (si NP : 99) **dont ganglions envahis N+** /__ __/
Si N+, précisez si gg lymphatiques régionaux homolatéraux /___/ non (0), oui (1), NP (9)
prétrachéaux / paratrachéaux / prélaryngés /___/ non (0), oui (1), NP (9)
autres chaînes cervicales ou médiastinales sup /___/ non (0), oui (1), NP (9)

2.8 pTNM **pT** /____/ **pN** /____/ **M** /____/ (si NP : 9) **numéro de l'édition utilisée** /____/

3 - DIAGNOSTIC FINAL [A]

/____/ **merci d'utiliser les n° de la rubrique « diagnostic final »**
2ème avis demandé /____/ non (0), oui (1) **si oui, diagnostic final posé après 2ème avis** /____/ non (0), oui (1)

4 - LES LESIONS ASSOCIEES [B]

/____/ /____/ /____/
NB si autre cancer thyroïdien associé (8) remplir : **diagnostic final du cancer associé** /____/

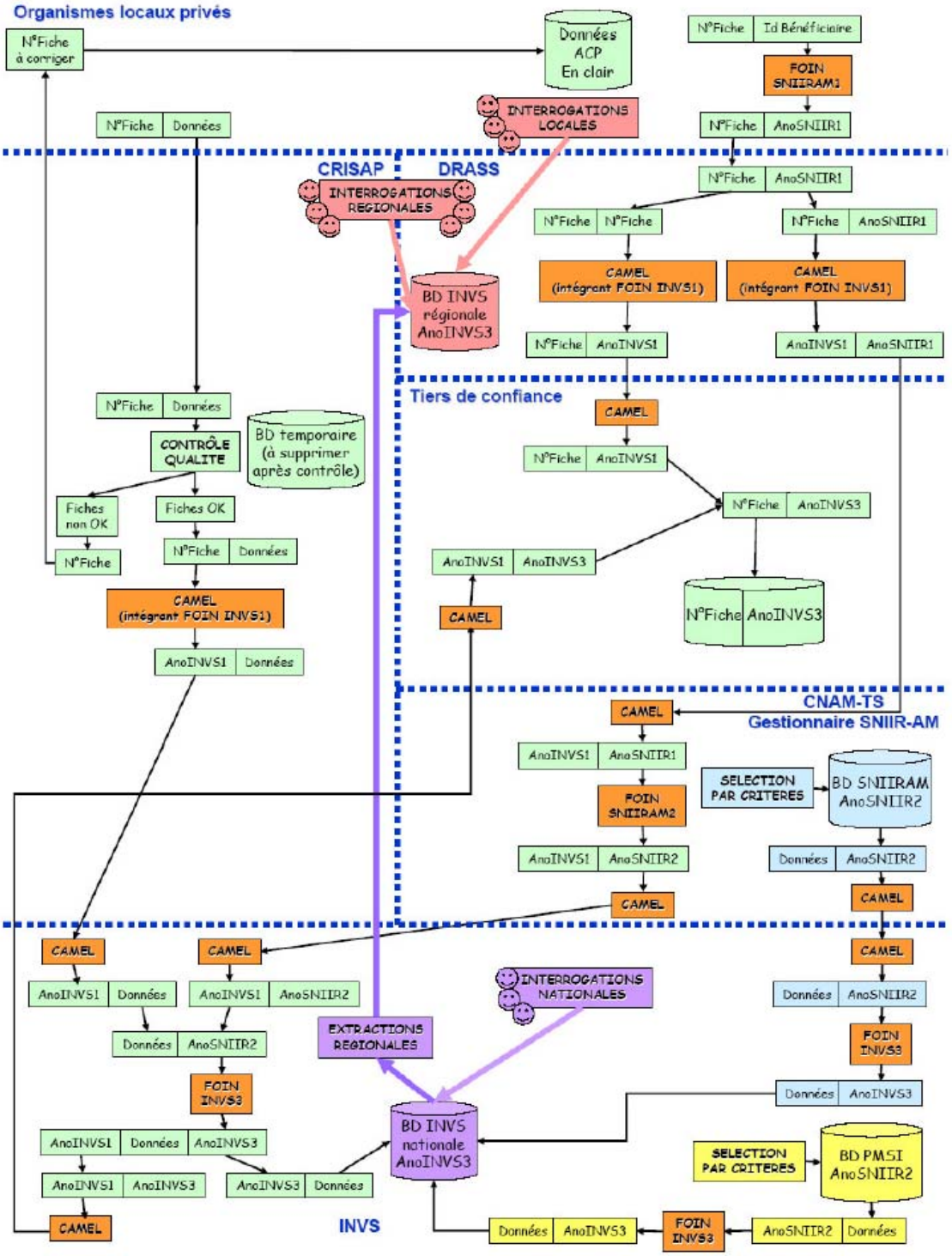
[A] **Diagnostic final** selon OMS et/ou AFIP (tous ces numéros seront transcodés automatiquement en ADICAP, SNOMED ou CIM10)

1.carcinome vésiculaire SAI
2.carcinome vésiculaire à invasion minime, encapsulé
3.carcinome vésiculaire largement invasif
4.carcinome vésiculaire à cellules oxyphiles, éosinophiles, oncocytaire, à cellules de Hürtle
5.carcinome vésiculaire à cellules claires
6. carcinome vésiculaire peu différencié et/ou insulaire
7.carcinome papillaire SAI
8.carcinome papillaire à variante vésiculaire
9.carcinome papillaire sclérosant diffus, occulte, non encapsulé
10.carcinome papillaire à cellules oxyphiles, éosinophiles
11.carcinome papillaire à cellules cylindriques, hautes
12.carcinome papillaire à présentation solide et/ou trabéculaire
13.carcinome médullaire SAI
14.carcinome mixte médullaire vésiculaire
15. carcinome mixte médullaire papillaire
16.carcinome médullaire héréditaire
17.carcinome anaplasique sans contingent bien différencié identifiable
18.carcinome anaplasique avec un contingent bien différencié minoritaire
19.carcinome anaplasique avec un contingent bien différencié majoritaire
20.carcinome indifférencié
21. carcinome SAI
22.lymphome
23.sarcome

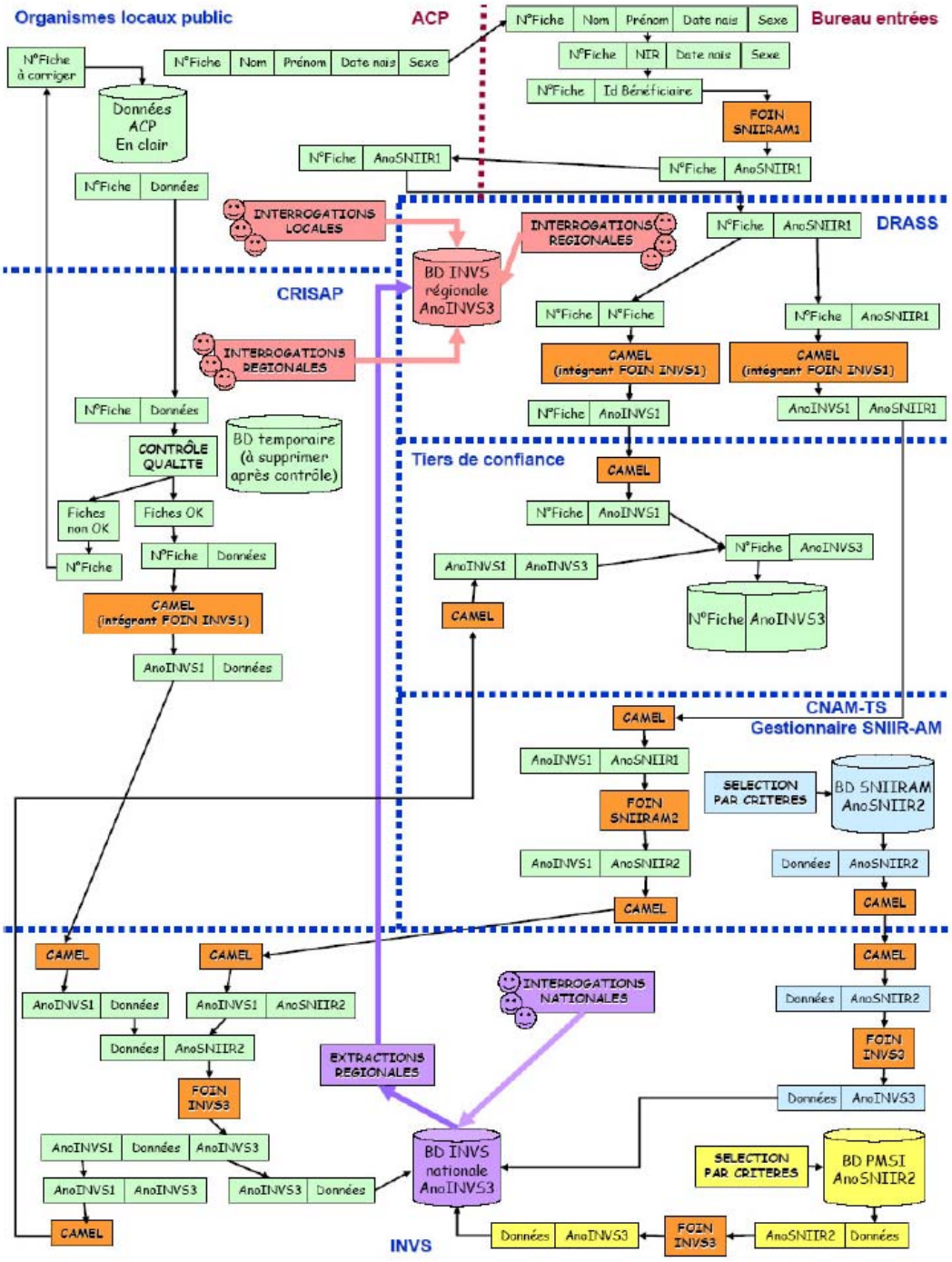
[B] **Lésions associées**

1.absence
2.hyperplasie diffuse ou nodulaire
3.thyroïdite lymphocytaire
4.adénome SAI
5.adénome atypique
6.adénome à cellules oxyphiles
7. adénome trabéculaire hyalinisant
8.autre cancer thyroïdien de type histologique différent
9.lésion fonctionnelle (Basedow, hyperplasie diffuse ou nodulaire)
10.nodule toxique

ANNEXE 3 – SCHÉMA D'ARCHITECTURE D'ANONYMISATION, STRUCTURES ACP PRIVÉES



ANNEXE 4 – SCHÉMA D'ARCHITECTURE D'ANONYMISATION, STRUCTURES ACP PUBLIQUES



ANNEXE 5 – LISTE DES ENTRETIENS REALISÉS AUPRÈS DES PARTENAIRES PAR L'AMOA (SOURCE : AMOA)

Structure	Fonction - Profession	Date entretien
ACP Privé 1	Médecin pathologiste	12/11/08
ACP Privé 2	Médecin pathologiste	09/12/08
ACP Public	Médecin pathologiste	10/12/08
Société savante ACP	Médecin pathologiste	21/11/08
Partenaire national	Responsable adjoint de service Chargé de projet	27/10/08
Partenaire national	Responsable de service Chef de projet	17/11/08
Partenaire régional	Chargé de projet Chargé de projet	20/11/08
Agence nationale	Responsable de département Chargé de mission	26/11/08
Société prestataire	Chef de projet	17/09/08

ANNEXE 6 – SYNTHESE DE LA PERCEPTION GÉNÉRALE DES PARTENAIRES INTERROGÉS (SOURCE : AMOA)

Les enjeux du projet SMSC

Enjeux globaux	- Recueil de données épidémiologiques de registre de cancer pour l'étude de la prévalence et l'incidence des cancers en vue d'alerter en cas de modification de l'incidence de ces cancers, mettre en évidence des foyers (au regard des facteurs environnementaux notamment) - Réalisation d'un relevé de données relatives aux cancers au niveau national - Système d'observation épidémiologique sur la prise en charge des patients atteints de cancer
Enjeux et apports spécifiques aux ACP	- Réalisation d'un relevé de données relatives aux cancers au niveau national - Rien d'identifié en terme de retour d'informations, essentiellement participation à une mission de santé publique. - Acquérir une certaine notoriété - Bénéficier auprès des éditeurs des négociations réalisées dans le cadre du SMSC par l'InVS pour leur propre dispositif. - Être dans une dynamique d'échange, de partenariat avec l'InVS - Opportunité de tester la génération automatique de CR standardisé
Enjeux et apports spécifiques à la CnamTS	- Permettre un enrichissement des données CnamTS par des données médicales - Enrichissement du Sniiram de données médicales : meilleure gestion du risque par l'Assurance Maladie

Perception générale du projet SMSC

Perceptions positives	- Enrichissement du Sniiram de données médicales : meilleure gestion du risque par l'Assurance Maladie - La mise en place de ce type de système est incontournable - Intérêt évident de constitution d'une base centralisée - Ambitieux et louable - Bonne relation avec l'InVS
Perceptions négatives	- Un système très lourd - Une communication sur le projet insuffisante - Ce projet ne peut être mis en œuvre que si les revendications sont levées (reconnaissance profession et revalorisation financière) - Pour les structures ACP, impossibilité de se désolidariser du mouvement national lancé par le syndicat.

Phase d'expérimentation : enjeux et perceptions

Enjeux de la phase d'expérimentation	- Vérifier la faisabilité technique à tous les niveaux du recueil des données / Soulever les questions posées par la complexité liée à l'anonymisation / Etude de faisabilité pour vérifier la viabilité technique du dispositif sur la base de données de 2004. - Nécessité d'ouvrir sur une phase d'expérimentation plus large - Tester le croisement des bases appartenant à des institutions différentes. - Identifier les difficultés rencontrées
Perception du déroulement de la phase d'expérimentation	- Commencée depuis 4 ans, cette phase paraît longue. Mais devrait aboutir. - Pas de difficulté particulière au niveau du temps consacré à l'installation (faite par le prestataire), la collecte des données administratives et la saisie des fiches - Interrogations sur l'objectif de cette phase dans la mesure où la faisabilité ne peut être réellement vérifiée puisque le dispositif expérimenté n'est pas le dispositif final (au niveau des structures ACP tout au moins). - Charge supplémentaire de récupération du NIR pour les fiches traitées auprès des cliniques. (peu de difficultés à l'obtenir) - Pas niveau d'effort très élevé, dans la moyenne, hormis le problème d'installation. Temps personnel important pour certains.

Phase de généralisation : enjeux et perceptions

Contraintes organisationnelles	- Ambiguïté quant à la faisabilité qui pourrait être conclue à l'issue de cette phase d'expérimentation dans la mesure où le dispositif n'est pas applicable sur le terrain.
Contraintes politiques	- Dans un contexte où le coût du travail des ACP est en augmentation constante et le rapport nombre de structures ACP/habitant diminue, les structures n'envisage pas de participer à un système de remontées d'informations générant une charge supplémentaire de travail sans compensation financière.
Contraintes techniques	- Pour la CnamTS, l'architecture d'anonymisation adoptée pour l'expérimentation ne peut être la même pour la généralisation
Contraintes juridiques	- Pour la CnamTS, la demande de copropriété des données sera renouvelée.

Facteurs clés de succès identifiés pour la réussite du projet SMSC

Communication	- Rôle primordial de la communication pour mener à terme un tel projet. ▪ Jugée insuffisante durant l'expérimentation ▪ Alternance de périodes de demandes de mobilisation intensive des équipes avec des périodes sans activité et sans informations quant à l'avancement - Des périodes sans évènement, communication qui donnaient l'impression que le projet était arrêté. - Communication insuffisante auprès des ACP pour lesquels le système SMSC est inconnu d'une manière générale. Pour ceux qui le connaissent, ce système est perçu comme un système espion qui viendrait se servir en données.
Une nécessité de prise en compte des contraintes des acteurs du projet	- Nécessité de disposer d'une solution entièrement automatisée (automatisation des remontées depuis les structures ACP) : une surcharge de travail supplémentaire semble très difficile à accepter. - Engagement des ACP à condition de participer à un projet construit et sérieux. - Effort souvent pris sur du temps personnel. Niveau d'effort dépend de la taille de la structure - Les ACP ont le sentiment de fournir des données sans contrepartie, sans écoute de leurs problèmes.
La nécessité d'une coordination des projets autour du cancer	- Meilleure coordination des acteurs « Cancer » et de leurs projets. Exemple : Coordination avec projets de l'INCa, de l'InVS et les éditeurs (Environ 15 éditeurs) pour une cohérence globale - Impression qu'il y a autant de dispositifs que de demandeurs - Impression des ACP que des registres, bases de données apparaissent à droite à gauche sans concertation des acteurs.

Autres informations intéressantes

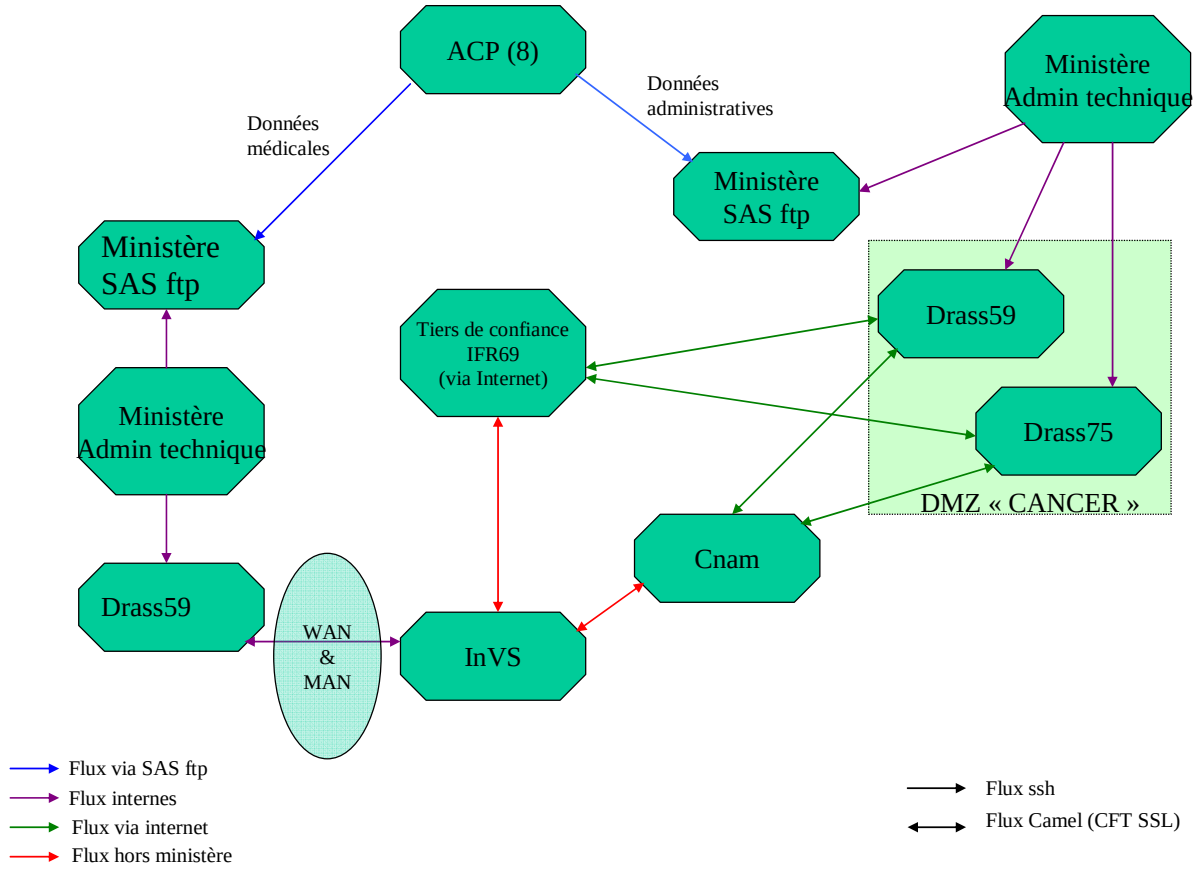
DCC et SMSC	- Le DCC pourrait être opérationnel plus rapidement que le DMP. - ACP non sollicités actuellement et ne savent pas où en est le projet - Distinction entre le DCC et le SMSC pas nécessairement perçue.
Utilisation des codes ADICAP	- L'interprétation des ces codes peut être différente d'un ACP à un autre (estimation à 10% d'erreur par rapport à l'analyse du diagnostic en clair) : un travail sur la qualité des codes ADICAP doit être mené pendant l'étude pilote. Jusqu'à présent, pas d'implication collective des ACP sur ces aspects : pas d'intérêt de ce codage du point de vue du patient. Par contre, ce codage est essentiel en termes de santé publique.
Relais régionaux de remontée / Pérennité des CRISAP	- Recommandation du Pr Belloch pour la mise en œuvre d'un système de relais des remontées au niveau régional (collecte et épuration des données avant la remontée à un niveau national), au sein d'une structure composée de différents partenaires impliqués dans la gestion des cancers. ▪ Ex : structure qui réunira différents partenaires ayant un lien avec le cancer, impliqués mais pas nécessairement responsables : Anapath, InVS et les Cires, médecins soignants, structures de gestion de dépistage, associations de patients, registres… et un professionnel des données - Pour le Pr Belloch, ce niveau ne peut être assuré par les CRISAP. Le CRISAP est une tentative d'ACP qui voulaient gérer leurs données pour les valoriser. A bout de 15 ans, seuls 2 CRISAP fonctionnent : Idf et PACA Est (ne comprend pas Marseille). Le constat est que c'est un échec.
Alternatives à la saisie complémentaire	- Une saisie complémentaire pour un nombre de cas limités peut être envisagée.
Communication au congrès des pathologistes	- L'InVS devra faire face aux craintes concernant le partage de données par des actions de communication, lors du congrès des pathologistes par exemple, pour démontrer le bien-fondé du projet.

ANNEXE 7 – PHASAGE DU PROJET

Blocs	Etapes	Description	Livrables	Délai de réalisation	Délais de validation par l'équipe projet
			PARTIE I		
Bloc 1	1	Réunion de lancement avec l'équipe projet	Compte rendu	1 semaine	1 semaine
	2	Entretiens SSI InVS et partenaires (structures ACP locales, CRISAPIF, Drass, CnamTS, IFR 69)	Spécifications fonctionnelles et techniques	6 semaines	4 semaines
	3	Fourniture par l'InVS des logiciels Foin et Camel			
Bloc 2	4	Développement, installation et configuration de la solution au niveau des trois structures ACP libérales privées (2 IdF + 1 NPdC), comprenant : - La récupération des données, - l'anonymisation (logiciel Foin), - la préparation de la transmission vers les deux partenaires du niveau régional, - l'automatisation de la procédure.	Application, sources et documentations	3 semaines	3 semaines
	5	Développement, installation et configuration de la solution au niveau des 2 Drass (IdF et NPdC) comprenant : - le serveur de réception sécurisé, - l'intégration de Foin et paramétrage du logiciel Camel, - la préparation de la transmission vers la CnamTS et l'IFR 69, - l'automatisation de la procédure.	Application, sources et documentations	2 semaines	2 semaines
	6	Développement, installation et configuration de la solution au niveau de cinq structures ACP hospitalièrers publiques (3 IdF + 2 NPdC), comprenant : - la récupération des données, - la récupération du NIR dans le SIH (connecteur SIH à mettre en place), - l'anonymisation (logiciel Foin), - la préparation de la transmission vers les deux partenaires du niveau régional, - l'automatisation de la procédure.	Application, sources et documentations	8 semaines	5 semaines
	7	Développement, installation et configuration de la solution au niveau du Crisapif et de la Drass Nord-Pas-de-Calais (équipe ACP) : - le serveur de réception sécurisé, - la base de données, - l'application de validation des données, - l'intégration et le paramétrage du logiciel Camel, - la préparation de la transmission vers l'InVS, - l'automatisation de la procédure.	Application, sources et documentations	3 semaines	1 semaine
Bloc 3	8	Préparation et automatisation de la transmission entre la CnamTS et l'InVS. *Pas d'intervention du prestataire à cette étape.*	Sans objet	1 semaine	1 semaine

Blocs	Etapes	Description	Livrables	Délai de réalisation	Délais de validation par l'équipe projet
PARTIE I					
	9	Développement, installation et configuration de la solution au niveau de l'InVS permettant pour les données ACP: - l'intégration et le paramétrage du logiciel Camel, - la gestion de l'appariement des flux ACP avant l'anonymisation, - l'intégration et le paramétrage du logiciel FOIN, - la mise en place de la base ORACLE, - le paramétrage de l'ETL pour importer les données ACP dans ORACLE, - la préparation de la transmission vers l'IFR 69.	Application, sources et documentations	3 semaines	2 semaines
	10	A l'IFR 69 : - Intégration et paramétrage de Camel pour la réception des données en provenance des Drass et de l'InVS ; mise en place d'une table de correspondance.	Application, sources et documentations	1 semaine	1 semaine
Bloc 4	11	A l'InVS : - paramétrage de Camel intégrant Foin (InVS3) pour les données ALD30 du Sniiram ; - paramétrage de l'ETL pour importer les données ALD30 dans ORACLE,	Application, sources et documentations	1 semaine	1 semaine
	12	A l'InVS : - mise en place de l'anonymisation de la base PMSI avec Foin (InVS3) ; - paramétrage de l'ETL pour importer les données PMSI dans ORACLE,	Application, sources et documentations	1 semaine	1 semaine
	13	Tests de croisement des 3 sources de données dans ORACLE	Application, sources et documentations	1 semaine	3 semaines
Bloc 5	14	Vérification de service régulier sur les 8 structures ACP locales à compter de la validation de l'étape 14.			3 mois
	15	Formation et transfert de compétences	Supports de formation	2 semaines	
PARTIE II					
Bloc 6	16	Garantie		8 mois	
Bloc 7	17	Maintenance		12 mois	

ANNEXE 8 – SCHÉMA DES FLUX



Système multi sources de surveillance des cancers (SMSC)

Etude pilote sur le cancer de la thyroïde en Ile-de-France et dans le Nord-Pas-de-Calais, 2010

Rapport final d'étude - Octobre 2010

Le développement d'un système de surveillance nationale, le Système multi sources de surveillance des cancers (SMSC), en complément du système reposant sur les registres a été recommandé par la commission d'évaluation des conséquences sanitaires de l'accident de Tchernobyl.

Le principe repose sur le croisement de données anonymisées de trois sources (PMSI, ALD et compte rendu ACP).

L'objectif de l'étude pilote était d'étudier la pertinence et la faisabilité de la mise en place du SMSC.

L'étude a porté sur le cancer de la thyroïde dans deux régions : Ile-de-France et Nord-Pas-de-Calais.

Le recueil des données du compte rendu ACP a été réalisé rétrospectivement (2004) grâce à un formulaire informatisé.

423 fiches de données ACP de cancer de la thyroïde ont été transmises par les ACP volontaires des deux régions. Elles ont été croisées avec celles des ALD cancer (1 159) et du PMSI (2 947) de l'année 2004.

Le taux d'appariement (nombre de cas ACP retrouvés dans les autres sources) observé était de 84 % (trois sources), de 81 % (ACP, PMSI) et de 48 % (ACP, ALD cancer).

La mise en place du système a mis en évidence que la solution informatique retenue était complexe pour un développement au niveau national. Cette complexité est liée aux difficultés techniques rencontrées au cours du projet.

A l'issue du pilote, deux axes de développement du SMSC au niveau national ont été identifiés :

- dans l'immédiat, croisement des deux sources PMSI et ALD cancer, à des fins descriptives et de production d'indicateurs de surveillance ;
- à moyen terme et en parallèle avec le premier axe, étudier les conditions de mise en place de la base nationale de données ACP.

Mots clés : cancer, surveillance, SMSC, données médico-administratives, ACP

Multi source surveillance system of cancers (MSSC): pilot study on thyroid cancer in Ile-de-France and Nord-Pas-de-Calais, 2010

The development of a national multi source surveillance system of cancers (MSSC) in addition to the registry-based system has been recommended by the Assessment Commission on Health Effects of the Chernobyl Accident.

The principle is to cross anonymous data of three sources (PMSI, ALD and pathologist's report called ACP).

The objective of the pilot study was to study the relevance and the feasibility of the MSSC implementation.

The study concerned thyroid cancer in two regions: Ile- de-France and Nord-Pas-de-Calais.

ACP data were retrospectively (2004) collected by electronic form.

Four hundred and twenty-three forms of ACP data on thyroid cancer were transmitted by pathologists of both regions on a voluntary basis. They were crossed with data on ALD thyroid cancer (1,159) and PMSI thyroid cancer (2,947) for the year 2004.

The rate of matched data (number of ACP cases retrieved in other sources) observed was 84% (3 sources), 81 % (PMSI, ACP), and 48% (ACP, ALD cancer).

The pilot showed that the implementation of the system at national level would be too complex due to technical problems related to computer systems.

At the end of the pilot, two lines of MSSC development at national level were identified:

- *In the short term, crossing both PMSI and ALD cancer sources, for descriptive purposes, and in order to produce surveillance indicators.*
- *In the mid term, and in parallel with the first line, study the conditions of ACP data implementation at national level.*

Citation suggérée :

Caserio-Schönemann C, Kudjawu Y, Chérié-Challine L, Guillet A, Musset A, Nicolau J *et al.* Système multi sources de surveillance des cancers (SMSC). Etude pilote sur le cancer de la thyroïde en Ile-de-France et dans le Nord-Pas-de-Calais. Rapport final d'étude, octobre 2010. Saint-Maurice: Institut de veille sanitaire; 2011. 68 p. Disponible à partir de l'URL : http://www.invs.sante.fr

INSTITUT DE VEILLE SANITAIRE
12 rue du Val d'Osne
94415 Saint-Maurice Cedex France
Tél. : 33 (0)1 41 79 67 00
Fax : 33 (0)1 41 79 67 67
www.invs.sante.fr

ISSN : 1956-6964
ISBN : 978-2-11-099498-1
ISBN NET : 978-2-11-128436-4
Tirage : 103 exemplaires
Impression : France-Repro
Dépôt légal : août 2011